

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:09:08 : Search time 968.083 Seconds  
(without alignments) 6775.784 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486  
Sequence: 1 atgacacagagatgttgcga.....aagtcattcacaagaaaaa 486

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.8	11.1	998	12	BH133989 ENTNM17TF
2	52.4	10.8	868	12	BH162606 ENTNV08TF
3	49.6	10.2	279	12	CNS03W9F
4	47.2	9.7	845	12	AZ685166 Tetracodon
5	46.4	9.5	325	12	CNS03PK8
6	46.4	9.5	624	10	BT001798
7	46.4	9.5	872	10	BT009355
8	46	9.5	909	12	CNS00UTL
9	46	9.5	939	12	CNS00CNG
10	44.6	9.2	890	12	BH164851
11	43.2	8.9	427	9	AU052962
12	43.2	8.9	440	9	AU052941
13	42.6	8.8	442	10	BE801262
14	42.6	8.8	452	12	BT471035
15	42.6	8.8	536	12	CNS0300K
16	42.2	8.7	748	10	BT176637
17	42.2	8.7	805	12	AZ753057

c 18	41.8	8.6	723	12	AZ627893
c 19	41.8	8.6	745	12	CNS01MHE
c 20	41.6	8.6	294	9	AU060898
c 21	41.2	8.5	1101	12	CNS00LOO
c 22	41	8.4	328	9	AI437626
c 23	41	8.4	344	10	BC352395
c 24	41	8.4	353	9	AI443457
c 25	41	8.4	402	9	AI442486
c 26	41	8.4	431	10	BE555999
c 27	41	8.4	496	10	BE805704
c 28	41	8.4	509	9	AW734942
c 29	41	8.4	520	10	BM093415
c 30	41	8.4	521	10	BE790034
c 31	41	8.4	536	10	BT178459
c 32	41	8.4	543	10	BT1701733
c 33	40.8	8.4	805	12	CNS04FW2
c 34	40.6	8.4	596	12	BH483310
c 35	40.6	8.4	676	10	BE603900
c 36	40.6	8.4	861	12	CNS0075A
c 37	40.4	8.3	173	10	BM159224
c 38	40.4	8.3	250	9	AU060486
c 39	40.4	8.3	444	9	AU033804
c 40	40.4	8.3	559	12	AZ304119
c 41	40.4	8.3	576	12	AZ379826
c 42	40.4	8.3	650	12	BH393056
c 43	40.2	8.3	1101	12	CNS00LT2
c 44	40	8.2	768	12	A0690676
c 45	40	8.2	1026	12	CNS003KS

#### ALIGNMENTS

RESULT 1	BH133989	998 bp	DNA	linear	GSS 07-AUG-2001
LOCUS	ENTNM17TF	Entamoeba histolytica	sheared DNA	Entamoeba histolytica	
DEFINITION	genomic DNA sequence.				
ACCESSION	BH133989				
VERSION	BH133989.1	GI:15093050			
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				
ORGANISM	Entamoeba histolytica.				
REFERENCE	1 (bases 1 to 998)				
AUTHORS	Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.				
TITLE	Determination of clone end sequences from Entamoeba histolytica				
JOURNAL	HMI:IMSS sheared DNA library (2001)				
COMMENT	Unpublished (2001)				
	Contact: Brendan J Loftus				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0208				
	Fax: 301 838 3543				
	Email: b.loftus@tigr.org				
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared				
	DNA library				
	Seq primer: M13-Forward				
	Class: Shotgun				
	High quality sequence start: 17				
	High quality sequence stop: 490.				
FEATURES	Location/Qualifiers				
source	1..998				
	/organism="Entamoeba histolytica"				
	/strain="HMI:IMSS"				
	/db_xref="taxon:5759"				
	/clone_lib="Entamoeba histolytica sheared DNA"				
	/note="Vector: pROSI; Site: 1. Bst I; Constructed at The				
	Institute for Genomic Research (TIGR), Rockville, MD.				
	Genomic DNA isolated from broth cultures of E. histolytica				
	using a method described by Clark and Diamond (Clark,				
	C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a				

268 9 200 2

Query Match	11.1%;	Score 53.8;	DB 12;	Length 998
Best Local Similarity	52.4%;			
Matches 118; Conserved				

[illegible][illegible]

**10**

981 TGTCTATGGTGGTGTTGCGTCGCTGTTGTTGCTGGTGTTGCAGCTGTTGGTA  
143

144 AGCATACTAATTGTTGACCAACTAACTCCGAAGAATCATGTTGTTGCCCTGTTGG

441 TGTGGTGTCTGTGTGGTGTGTCTGTGTGGTGTGATTAATTGCTTC 203

501 TACTGATGATAAAAGCTAA 248

.....GAACTGAAAA 545

Accession number	Gene	Size (bp)	Accession number	Gene	Size (bp)
BH162506	Entamoeba histolytica	868	BH162506	Entamoeba histolytica	868
ENTRV087F	Entamoeba histolytica	868	ENTRV087F	Entamoeba histolytica	868

	BH162606	GI::15736044	Echinamoeba histolytica
	BH162606.1	GSS	
WORDS			
SION			
..			

ORGANISM histolytica.  
Entamoeba histolytica.  
Eukaryota: Entamoeba

Loftus, B., Wang, Z., Van Aken, S. and Fraser C  
Determination of clove

Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Microbiology  
University of Liverpool  
L69 3GB  
United Kingdom  
Tel: +44 (0)151 794 3444  
Fax: +44 (0)151 794 3444  
E-mail: b.loftus@liverpool.ac.uk

The Institute for Genomic Genomics  
9712 Medical Center Drive, Rockville, MD 20850  
Tel.: 301 592 1000

Fax: 501 838 3543  
 Email: [bjloftus@tigr.org](mailto:bjloftus@tigr.org)  
 Clones are available

	Seq primer: M13-Forward	
	Class: shotgun	
		histolytica HM1:IMSS sheared

High quality sequence start: 30  
Location/Quality stop: 768.

```

/organism="Entamoeba histolytica"
/strain="HM1:IMSc"

```

```

/c/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1. c14"

```

Genomic DNA isolated from broth cultures of *E. histolytica* was constructed at The Genomic Research (TIGR), Rockville, MD, using a method described by Sambrook et al. (1989).

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb), the v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Maizumi, B. Barrell, Oxford University Press, 1999).

On the other hand,

Best Local Similarity	Score	DB	Length
Matches 104; Conservative 0	52.4;	12;	868;
	54.7%;		
	Pred. No. 0.0012;		

[illegible]

84 AGAGTCGCGAGCTGCAGACTGGGCGCTGTGTGCCCGAAT  
TCCACGACGATGGTGCTGTTGCTGCTGTTGTCAATTGCTGTGT

144 AGCTAGTGGTGCACCTCC... 172

172 TGTGGTGCCTCTGTTGGTGCTGCTGTTGGTGGTAAATTCCTC 203

112 TGTGTGGT 103

SULT 3  
S03W9F/C

DEFINITION	279 bp	DNA	linear	GSS 18-may 2001
cas05w9f				
Tetradodon nigroviridis genome survey contig				
063D14 of library				

AL263436  
AL263436.1  
GI:7985000

Tetraodon nigroviridis.  
Tetraodon nigroviridis  
Tetraodon nigroviridis

[illegible]

2. (bases 1 to 279)  
Roest-Crolius, H., Tallon, O., Pasitua  
Bonneau, I., Bittner, L.

Characterization and repeat analysis of the freshwater bacteriophage  $\phi$ 229E. Bernot, A. and Saurin, W., *Journal of Virology*, 1990, 64, 10, 4102-4106.

2 (bases 1 to 279)  
Roest-Crolius/H. J. Talignon

Saurin, W. and Weissensbach, J., Brotlier, P., Quetier, F., Bouneau, L., Fisher, C., Human gene number estimate provided by tetraded analysis of the human genome.

Genoscope  
3 (bases 1 to 279)  
unpublished  
sequence  
analysis using

Submitted (12-APR-2000) to the EMRI/Coordination Unit  
This sequence is a direct submission

genome. For more information, please take a look at <http://www.genoscope.com> and was generated as part of a large tetraodon nigroviridis

1. .279  
/organization="Not a  
location/qualifiers

[/db\\_xref="taxon:99883" /clone="063D14"](#)

```
/clone_lib="G"
/Note="Genoscope sequence ID : C0BG063DB07SP1-end :
PUC-ori"
BASE COUNT      106 a      155 c      2 g      13 t      3 others
ORIGIN

Query Match
Best Local Similarity 56.8%; Pred. No. 0.0052;
Matches 88; Conservative 1; Mismatches 66; Indels 0; Gaps 0;

Qy 47 TGTATTAGTGGCTGCTGCCAAGACTTATGCTCAGACGAGCGACGCTGGGG 106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 TGTGCGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 160
Qy 107 CTGTTGTCGCGCTGTCGCGCAGCTGTTGGTAAAGTAGTGCAGTTCGAATGC 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
Qy 167 CCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 ATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65

RESULT 4
A2685166/c 845 bp DNA linear GSS 14-DEC-2000
LOCUS ENT1GL1TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION A2685166
VERSION A2685166.1 GI:11822312
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 845)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
CONTACT: Brendan J Loftus
DEPARTMENT OF Eukaryotic Genomics
THE INSTITUTE FOR GENOMIC RESEARCH
9712 MEDICAL CENTER DR., ROCKVILLE, MD 20850, USA
TEL: 301 838 0208
FAX: 301 838 3543
EMAIL: b.loftus@lifgr.org
CLONES ARE DERIVED FROM THE Entamoeba histolytica HMI:IMSS sheared
DNA library
SEQ PRIMER: M13-Forward
CLASS: shotgun
HIGH QUALITY SEQUENCE START: 26
HIGH QUALITY SEQUENCE STOP: 726.
LOCATION/Qualifiers
1. 845
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/Note="Vector: pHOSt1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Bartell, Oxford University Press, 1999)."
```

```
Query Match
Best Local Similarity 52.6%; Pred. No. 0.032;
Matches 103; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 74 TTAGTCGTCAAGAGTCGACGCTGCGAGCTGTTGGCGGTGCTGCGCAGC 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 TTGTCCTCTTTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 684
Qy 134 TCTTTGTTAAGTAGTGTGTCGACATTCGATGCGCATTTGCTGTTGGTGAT 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 TTGTCGTCCTGTTGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 624
Qy 194 TAAFTGCTTCAAAATGCGCATTCGATGATGATGATGATGATGATGATGATGATG 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 TTGTCGTCCTGTTGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 564
Qy 254 GTTTGCAAAAGTAAA 269
Db 563 TTGAAATTAATTAATAA 548

RESULT 5
CNS03PK8 325 bp DNA linear GSS 17-MAY-2000
LOCUS CNS03PK8/c
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
045118 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL254753
VERSION AL254753.1 GI:7975765
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
REFERENCE 1 (bases 1 to 325)
AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Fizesse,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 325)
AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizesse,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 325)
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
LOCATION/Qualifiers
1. 325
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="045118"
/clone_lib="G"
/Note="Genoscope sequence ID : C0BG045BE09SP1-end :
PUC-ori"
BASE COUNT      157 a      157 c      10 g      0 t      1 others
ORIGIN

Query Match
Best Local Similarity 52.6%; Pred. No. 0.04;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
```

```

OY 22 GGTAGTACTTAAATTATATCAGTGTGTTTAACTGGCTGTCGCCAGAACTTACGT 81
DB 237 GGTGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 178
OY 82 CAGAAGTCGAGCGCAGCAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 141
DB 177 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 118
OY 142 AAAGTAGTGTGCTGAGTTTGCATGAGCCATTGGTGGCGGTGGCGGCGAGCTGTTTGT 141
DB 117 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 201
OY 202 TCTAAATCGGT 213
DB 57 GTTGTGTGTGTGT 46

RESULT 6
BU001798/c 624 bp mRNA linear EST 05-DEC-2001
LOCUS BU001798 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA027D02 5'
DEFINITION BU001798 mRNA sequence.
ACCESSION BU001798
VERSION BU001798
KEYWORDS BU001798.1 GI:17364689
SOURCE EST.
ORGANISM Japanese medaka.
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Belontiiformes; Acanthopterygii; Percormorpha; Atherinomorpha;
1 (bases 1 to 624)
Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
LOCATION/Qualifiers
1. 624
/organism="Oryzias latipes"
/db_xref="taxon:8090"
/clone="MF01SSA027D02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"

BASE COUNT 208 a 248 c 64 g 104 t

Query Match
Best Local Similarity 9.5%; Score 46.4; DB 10; Length 624;
Matches 104; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

OY 14 GTTTCAGAGTAGTACTTAAATTATATCAGTGTGTTTAACTGGCTGTCGCCAGAACT 73
DB 453 GTTTCAGAGTAGTACTTAAATTATATCAGTGTGTTTAACTGGCTGTCGCCAGAACT 73
OY 74 TTAAGTCGTAAGAAGTCGAGCTGCGACTGGGCGGTGTTGGCGGTGCTGGCCAGC 133
DB 393 TTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 133
OY 134 TGTTCGTAAGAAGTCGAGTGTGCAATGAGCCATTGGTGGCGGTGCTGGCCAGC 334
DB 333 TAGACGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 193
OY 194 TAATGTGTCTTAAATCGGT 213

```

```

DB 273 CTGGTGTACTATAGACGT 254

RESULT 7
BU009355/c 872 bp mRNA linear EST 05-DEC-2001
LOCUS BU009355 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA131D04 5'
DEFINITION BU009355 mRNA sequence.
ACCESSION BU009355
VERSION BU009355
KEYWORDS BU009355.1 GI:17356777
SOURCE EST.
ORGANISM Japanese medaka.
Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Belontiiformes; Acanthopterygii; Percormorpha; Atherinomorpha;
1 (bases 1 to 872)
Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
LOCATION/Qualifiers
1. 872
/organism="Oryzias latipes"
/db_xref="taxon:8090"
/clone="MF01SSA131D04"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"

BASE COUNT 250 a 318 c 130 g 169 t 5 others

Query Match
Best Local Similarity 9.5%; Score 46.4; DB 10; Length 872;
Matches 104; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

OY 14 GTTTCAGAGTAGTACTTAAATTATATCAGTGTGTTTAACTGGCTGTCGCCAGAACT 73
DB 762 GTTTCAGAGTAGTACTTAAATTATATCAGTGTGTTTAACTGGCTGTCGCCAGAACT 73
OY 74 TTAAGTCGTAAGAAGTCGAGCTGCGACTGGGCGGTGTTGGCGGTGCTGGCCAGC 133
DB 702 TTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 133
OY 134 TGTTCGTAAGAAGTCGAGTGTGCAATGAGCCATTGGTGGCGGTGCTGGCCAGC 643
DB 642 TAGACGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 193
OY 194 TAATGTGTCTTAAATCGGT 213
DB 582 CTGGTGTACTATAGACGT 563

RESULT 8
CNS00JTL 909 bp DNA linear GSS 03-JUN-1999
LOCUS BAC33608 of RPI-98 library from Drosophila melanogaster (fruit
DEFINITION fly), genomic survey sequence.
ACCESSION AL076720
VERSION AL076720.1 GI:4956298
KEYWORDS GSS.
SOURCE fruit fly.

```





Query Match	8.98;	Score 43.2;	DB 9;	Length 427;
Best Local Similarity	51.68;			
Matches 99.				

```

/strain="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_id="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1. Site1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
tight size distribution for DNA identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."
```

Query Match	9.28;	Score	44.6;	DB	12;	Length	890;
Best Local	Similarity	65.78;					
Matches	65;	Conservative	0.17;				

[illegible][illegible]

b

656 GGTGCTGTTGGTGGTGCTGATTATTCGT 201  
|| ||| |  
c TCGCAGTTCGCTGCATTAATTCGT 201  
|| ||| |  
d CACCTGCTGCTGCTGCTGCTGCTCATTCGT 694  
|||||

RESULT 11  
U052962/C  
OCUS  
A0000000

[illegible][illegible]

ORGANISM

Dictyostelium discoideum.  
Dictyostelium discoideum.  
Eukaryota: Murogota

1 (bases 1 to 427); Dictyostelium.  
Morio, T., Ushihara, H., Saito, T.,  
Yoshino, K., and Iwano, N.

WILLIAMS, J., MITRA, B. N., PI, M., SATO, T., OGAWA, Y., MIZUNO, H., YOSHIDA, M.,  
MEEDA, M., TAKEUCHI, I., TAKEMOTO, K., YASUKAWA, H.,  
OCHIAI, I. and TANAKA, Y.,  
Developmental cDNA in *Dictyostelium discoideum*  
Unpublished (1998)

contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3

3-3-10 Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp  
PROJECT = Dictyostelium

Location/Qualifiers  
1. .427  
CDNA project in Japan.  
discoidium  
fasciculum  
qualifiers

```

/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"

```

```
clone="SLF389"  
/clone_lib="Dictyostelium discoideum SL (H Fruehling)"  
/dev_stage="slugs"  
COUNT 221
```

IN	231 a	105 c	25 g	66 f	( <i>Arctostaphylos</i> )

[illegible]

RESULT 12  
AU052941/c  
LOCUS  
AU052941

ACCESSION	DEFINITION	EST
AU052941	Dictyostelium discoideum SL (H. Drushl哈拉)	28-APR-1991
discoidium	CDNA clone SLF35	28-APR-1991
AU052941	Dictyostelium discoideum SL (H. Drushl哈拉)	28-APR-1991

```

VERSION      A0052941.1
KEYWORDS     GI:4701424
SOURCE       Dictyostelium

```

ORGANISM  
Dictyostelium discoideum.  
Dictyostelium discoideum  
Eukaryota; Mycetozoa. Dictyostelium

Yoshino, R., Morio, T., Urushihara, H., Saito, T., Uqawa, Y., Miyazaki, T. & I (bases 1 to 440) Dictyostelium.

TITLE	Author
Williams, J., Maeda, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.	Developmental CDNA in Dictyostelium
Unpublished (1990)	Ochiai, H. and Tanaka, Y.

COMMENT: .....  
Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
Tsukuba, Ibaraki 305-8572, Japan  
E-mail: urushihara@biology.tsukuba.ac.jp

3-3-10 Ten-nodal, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp

LOCUS = Dictyostelium discoideum cDNA project in Japan.  
location/Qualifiers  
1. .440

```
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db_xref="rayon:44000
```

```

/clone="SLF355"
/clone_lib="Dictyostellum discoideum SL (H.Trushihara)"
/dev_stage="slug"
110
234 a

```

Temperature	Weight	Time
440 C	26 g	70 t

Test Local Similarity	8.98;	Score 43.2;	DB 9;	Length 440;
atches	51.68;			
99; Conservative	0;	Pred. No. 0.33;		
		Mismatch		

	Indels	Gaps
11 GATGTTTCAGGTAGTACTTAATTATTATTCAGTGTCTTTAGTGGCTGCTGCCCTA	93	0
370 GATCT	0	0

[illegible]

310 TTGTTGTCTGTTGTCCTGTTCTGTGTTGGCGGTGTGCYGCC 130

**C**

131 AGCGTTTGGTAAGGACTGTGCACATTCATGCCCATTGCTGTGGCg	251
250 TTGTTGTGTCTCCctcccccgcacccaatcgttggcgcttttgagcg	190

191 GATTATATGCTT 202

```

Db      190 TTGTTGTTGTTT 179

RESULT 13
BE801262/c 442 bp mRNA linear EST 06-DEC-2001
LOCUS      srt1106.y1 Gm-c1050 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-c1050-83 5' similar to TR:Q40786 Q40786 ARABINOGALACTAN-PROTEIN
PRECUSOR   ; mRNA sequence.

ACCESSION  BE801262.1 GI:10232374
VERSION     BE801262
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
             Glycine.
REFERENCE  1 (bases 1 to 442)
AUTHORS    Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
            A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            R., Ritzer,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            R., Waterston,R. and Wilson,R.
            Public Soybean EST Project
            Unpublished (1999)
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Resgen, Invitrogen Corp. 2130
            South Memorial Parkway Hunttsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 893 Std Error: 0.00
            High quality sequence stop: 420.
            Location/Qualifiers
                1..442
                /organism="Glycine max"
                /db_xref="taxon:3847"
                /clone="GENOME SYSTEMS CLONE ID: Gm-c1050-83"
                /clone_1lb="Gm-c1050"
                /tissue_type="leaf tissue at various developmental stages
                of greenhouse grown plants"
                /dev_stage="3 week old"
                /lab_host="DH10B"
                /note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                XhoI; The Clark NIL was constructed and seed was provided
                by Dr. J. Specht, University of Nebraska (Shoemaker and
                Specht, 1995). The cDNA library was constructed from mRNA
                isolated from leaf tissue at various developmental stages
                of 3 week old greenhouse grown plants. Complementary DNA
                was synthesized from mRNA using a primer consisting of a
                poly(dT) sequence with a XhoI restriction site and a 3'
                anchor. EcoRI adapters were ligated to the blunt-ended
                cDNA fragments followed by XhoI digestion. The cDNA
                fragments were directionally cloned into the EcoRI-XhoI
                restriction site of the pBluescript vector. The ligated
                cDNA fragments were transformed into DH10B host cells
                (GibcoBRL). The library was constructed in cooperation
                with Dr. Paul Keim's laboratory at Northern Arizona
                University."

BASE COUNT  107 a 195 c 71 g 69 t

ORIGIN
Query Match      8.88; Score 42.6; DB 10; Length 442;
Best Local Similarity 53.38; Pred. No. 0.48;
Matches 90; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Db      116 ATCTTGTTGTTGTTGATGGGTTGTAGTGGTGTGTGCTG 68

RESULT 14
BI471035/c 452 bp mRNA linear EST 29-NOV-2001
LOCUS      sah92f06.y1 Gm-c1050 glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-c1050-3707 5' similar to TR:Q40786 Q40786
            ARABINOGALACTAN-PROTEIN PRECURSOR. ; mRNA sequence.
ACCESSION  BI471035.1 GI:15287144
VERSION     BI471035
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
             Glycine.
REFERENCE  1 (bases 1 to 452)
AUTHORS    Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
            A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            R., Ritzer,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            R., Waterston,R. and Wilson,R.
            Public Soybean EST Project
            Unpublished (1999)
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Resgen, Invitrogen Corp. 2130
            South Memorial Parkway Hunttsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 423.
            Location/Qualifiers
                1..452
                /organism="Glycine max"
                /db_xref="taxon:3847"
                /clone="GENOME SYSTEMS CLONE ID: Gm-c1050-3707"
                /clone_1lb="Gm-c1050"
                /tissue_type="leaf tissue at various developmental stages
                of greenhouse grown plants"
                /dev_stage="3 week old"
                /lab_host="DH10B"
                /note="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                XhoI; The Clark NIL was constructed and seed was provided
                by Dr. J. Specht, University of Nebraska (Shoemaker and
                Specht, 1995). The cDNA library was constructed from mRNA
                isolated from leaf tissue at various developmental stages
                of 3 week old greenhouse grown plants. Complementary DNA
                was synthesized from mRNA using a primer consisting of a
                poly(dT) sequence with a XhoI restriction site and a 3'
                anchor. EcoRI adapters were ligated to the blunt-ended
                cDNA fragments followed by XhoI digestion. The cDNA
                fragments were directionally cloned into the EcoRI-XhoI
                restriction site of the pBluescript vector. The ligated
                cDNA fragments were transformed into DH10B host cells
                (GibcoBRL). The library was constructed in cooperation
                with Dr. Paul Keim's laboratory at Northern Arizona
                University."

FEATURES
source

```

BASE COUNT 111 a 193 c 67 g 81 t  
 ORIGIN

Query Match  
 Best Local Similarity 8.8%; Score 42.6; DB 10; Length 452;  
 Matches 90; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 35 TTATATACAGTGTGTTTATGTCGTCGCCAGAACTTTAGCTCAAGAAAGTCGAG 94  
 Db 304 TTGCTGTGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 94  
 QY 95 CTGCGACTGGGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 245  
 Db 244 CTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 154  
 QY 155 GAGTTGCAATGCCATTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 185  
 Db 184 ATCTTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 136

RESULT 15  
 CDS0300K/c  
 LOCUS  
 DEFINITION Tetraodon nigroviridis 536 bp DNA linear GSS 17-MAY-2000  
 046624 of library G from Tetraodon nigroviridis, genomic survey

ACCESSION AL255341  
 VERSION AL255341.1 GI:7976353  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Tetraodontidae; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 1 (bases 1 to 536)  
 Roest-Crolius H., Jallion O., Dasilva C., Fizes C., Fisher C.,  
 Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
 Weissenbach V.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 2 (bases 1 to 536)  
 Roest-Crolius H., Jallion O., Dasilva C., Fizes C., Fisher C.,  
 Saurin W., Fizes C., Wincker P., Brothier P., Quetier F.,  
 Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence  
 3 (bases 1 to 536)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.  
 Location/Qualifiers  
 1. 536  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="046624"  
 /clone\_1lb="G"  
 /note="Genoscope sequence ID : C08G046BD12LP1-end : 77"

BASE COUNT 163 a 236 c 79 g 43 t 15 others  
 ORIGIN

Query Match  
 Best Local Similarity 8.8%; Score 42.6; DB 12; Length 536;  
 Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 77 GTCTCAAGAGTGGAGCTGCGCTGTGTGCGCTGTGTGCGCTGTGCGCTGTGCTGTG 136

Db 423 GTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1  
 QY 137 TTGCTAAAGTATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 364  
 Db 363 GTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 196  
 QY 197 TTGCTTAAATGCTGT 213  
 Db 303 TTGCTGTGCTGTGCTGTGCT 287

Search completed: October 27, 2002, 18:34:51  
 Job time : 973.083 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 16:06:08 : Search time 35.5337 Seconds  
(without alignments)  
3359.570 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486  
Sequence: 1 atgaacagcggatgttgc.....aagtcattcacaagaanaaa 486

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued\_Patents\_NA:\*  
1: /cgn2.6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2.6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2.6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2.6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2.6/prodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2.6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.8	9.8	7218	1	US-08-232-463-14
C 2	42.6	8.8	465	4	US-08-476-102A-2
C 3	35.8	7.4	303	4	US-08-556-978B-80
C 4	35.8	7.4	303	4	US-08-556-978B-81
C 5	35.4	7.3	303	4	US-08-556-978B-82
C 6	35.2	7.2	198	5	PCT-US95-10668-3
C 7	34.4	7.2	198	5	PCT-US95-10668-4
C 8	34.4	7.1	1046	1	US-08-361-467B-4
C 9	34.4	7.1	1046	1	US-08-484-332C-4
C 10	33.6	6.9	198	5	PCT-US95-10668-1
C 11	33.6	6.9	198	5	PCT-US95-10668-2
C 12	33.2	6.8	1011	2	US-08-825-781-2
C 13	33.2	6.8	2072	2	US-09-073-362-2
C 14	33.2	6.8	2072	2	US-09-243-920-2
C 15	33.2	6.8	377	2	US-08-332-765A-1
C 16	32.8	6.7	1690	2	US-08-276-452A-24
C 17	32.8	6.7	1690	2	US-08-798-744-24
C 18	32.4	6.7	1984	1	US-07-885-970A-25
C 19	32.4	6.7	1985	1	US-08-298-687A-25
C 20	32.4	6.7	1985	1	US-08-298-829-25
C 21	32.2	6.6	1561	1	US-07-968-971A-3
C 22	32.2	6.6	1561	1	US-07-824-247-44
C 23	32.2	6.6	1561	1	US-08-142-473A-4
C 24	32.2	6.6	1561	1	US-08-469-203A-4
C 25	32.2	6.6	1561	1	US-08-469-203A-4
C 26	32.2	6.6	1561	3	US-08-470-204A-44
C 27	31.8	6.5	379	1	US-08-145-617-5

28	31.8	6.5	2793	1	US-08-209-747-1	Sequence 1, Appl1
29	31.8	6.5	2793	1	US-08-458-298-1	Sequence 1, Appl1
30	31.6	6.5	5496	4	US-09-462-284-1	Sequence 1, Appl1
C 31	31.6	6.5	1632	1	US-08-324-243-34	Sequence 34, Appl1
C 32	31.6	6.5	1632	1	US-08-532-390-34	Sequence 34, Appl1
C 33	31.6	6.5	1632	1	US-08-717-294-34	Sequence 34, Appl1
C 34	31.6	6.5	1632	5	PCT-US95-11511-34	Sequence 34, Appl1
C 35	31.4	6.5	925	3	US-08-858-003-1	Sequence 1, Appl1
C 36	31.4	6.5	925	3	US-09-078-166-1	Sequence 1, Appl1
C 37	31.4	6.5	925	4	US-08-997-467-1	Sequence 1, Appl1
C 38	31.4	6.5	3254	4	US-08-965-903B-1	Sequence 1, Appl1
C 39	31.4	6.5	3381	4	US-09-336-447A-6	Sequence 6, Appl1
40	31.2	6.4	289	4	US-09-007-005-17	Sequence 17, Appl1
41	31.2	6.4	289	4	US-09-244-796-17	Sequence 17, Appl1
C 42	31.2	6.4	744	4	US-08-936-165A-75	Sequence 75, Appl1
C 43	31.2	6.4	1086	1	US-08-415-751-47	Sequence 47, Appl1
44	31.2	6.4	2242	1	US-08-641-627A-37	Sequence 37, Appl1
C 45	31	6.4	1897	1	US-08-453-477-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
CLASSIFICATION: 435  
FILING DATE:  
PRIOR APPLICATION DATA:  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pC-Fts  
US-08-232-463-14  
Query Match 9.8%; Score 47.8; DB 1; Length 7218;



```

US-08-556-978B-81
: Sequence 81, Application US/08556978B
: Patent No. 6268169
: GENERAL INFORMATION:
: APPLICANT: FAHNESTOCK, STEPHEN F.
: TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
: TITLE OF INVENTION: SPIDER SILK ANALOGS
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
: STREET: 1007 MARKET STREET
: CITY: WILMINTON
: STATE: DELAWARE
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 19898
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.50 INCH
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: MICROSOFT WINDOWS 95
: SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/556, 978B
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/077, 600
: FILING DATE: JUNE 15, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: FLOYD, LINDA AXAMETHY
: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: CR-9389-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 302-892-8112
: TELEFAX: 302-773-0164
: INFORMATION FOR SEQ ID NO: 81:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 303 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-556-978B-81

Query Match 7.4%, Score 35.8; DB 4; Length 303;
Best Local Similarity 56.3%; Pred. No. 0.032;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 84 AGAAGTCGAGACTGGCACTGGGGGCTGTGTGGCGCGCTGTGCTGGCCACACTTTTGGTAA 143
Db 83 AGCCCTCGCGGACAGCTGGTGTGGTGCAGGGCGGCTGTGGCTCAACAAGGGGCGGCTCA 142
OY 144 AGGTAGTCGTCGACAGTGGCAATGGCCATTTGGTGTGCTGTGGGCGGATTAATGGTT 202
Db 143 AGGCCCTGCTGCAGCAGCAGCAGCTGCCGCTGGCGGTGCAGAGCCAAAGGTGATAGTGCT 201

RESULT 5
US-08-556-978B-82
: Sequence 82, Application US/08556978B
: Patent No. 6268169
: GENERAL INFORMATION:
: APPLICANT: FAHNESTOCK, STEPHEN F.
: TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
: TITLE OF INVENTION: SPIDER SILK ANALOGS
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
: STREET: 1007 MARKET STREET
: CITY: WILMINTON
: STATE: DELAWARE
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 19898
: COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556, 978B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077, 600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-556-978B-82

Query Match
Best local similarity 7.3%; Score 35.4; DB 4; Length 303;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0

QY 62 GTGCCAGAACTTATGTCGTCAGAAGTTCGGAGCTGCGACCTGGGGCTGTTGTGGCGGTG 121
    ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 GTGGCCAGGATGACAGTGTCTGCTGCTGCGGCTGTGTGTGTCAGGTCAAGGATGATC 127
QY 122 TTGCTGCGCCAGCTGTTTGGTAAGGTAGTGTGTCAGATTGCATGGCCATTGGTGGTCTG 181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 128 TGGGATCTCAGAGGGCCAGAGTCAGAGTGTGTGTGTCAGAGCTGCGGCGACGTGTGGCGCG 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 TTTTGGTGTGATTAATTGATT 202
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 188 GTCAAGGTGCTACGCGCGGTT 208
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
PCT-US95-10668-3/C
Sequence 3, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294, 133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:

```

us-09-677-374-1.rtf

Query Match	7.28;	Score 35.2;	DB 5;	Length 198;
Best Local Similarity	49.58;	Pred. No. 0.04;		
Matches 91;	Conservative	0.000000		

RESULT 7  
PCT-US95.

10

Query Match	7.28;	Score 35.2;	DB 5;	Length 198;
Best Local Similarity	49.58;			
Matches	91; Conservatio	Pred. No. 0.04;		

RESULT 8  
US-08-36

TELEFAX: (703) 836-2021



```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1046 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IMMEDIATE SOURCE:
CLONE: 3C9
US-08-484-332C-4

Query Match 7.1%; Score 34.4; DB 1; Length 1046;
Best Local Similarity 52.9%; Pred. No. 0.18;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 84 AGAAGTCGACCTCGACTGGCGCTGTTGGCGCGCTTCTGCGCAGCTGTTGCTAA 143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 AGGAGCAGCATTAAGATGATGATGTTGGTGTGCTTAACTGCTAGCTGCGTGG 270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 144 AGGATGCTGAGTGCATTAAGCCATGGCGATGGCGTCTTTGGGTGATTAATTGCTC 203
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 TTTTGATGCGCTATTAAGTGGAGCGCTTATGTTGGTGCTTAACCGGTGTTAACTTGCG 210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 204 TAAATCGGTCATCGATG 223
      | ||| ||| |||
Db 209 TTTGATGCGCTATTAAGTGG 190
      ||| ||| ||| |||

RESULT 10
PCT-US95-10668-1/c
Sequence 1, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: NO

```





Tue Oct 29 08:30:52 2002

us-09-677-374-1.rni

Page 8

ADDRESSEE: CUSHMAN DAREY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,766A  
FILING DATE: 01-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9326052.9  
FILING DATE: 21-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BIRD, Donald J.  
REGISTRATION NUMBER: 25,323  
REFERENCE/DOCKET NUMBER: 21/211/M94/0434/GB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
INFORMATION FOR SEQ. ID NO.: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-332-766A-1

Query Match  
Best Local Similarity 6.8%; Score 33; DB 2; Length 377;  
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 78 TCGTCAAGAGTGGAGCTGCGACTGCGGCTGTGTGGCGGCTGTGCGCCAGCTGT 137  
DB 66 TCCTCCTGATACAAATGGTATGATGATGATGATGATGATGATGATGATGATGATG 137  
QY 138 TGGTAAAGTAGTGTGCGAGTTGCAATGCCATTGSGTGTGCTGCTGTTGGGGATTAAT 197  
DB 126 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 125  
QY 196 TGGTCTCTAA 206  
DB 186 TGTGTGTGA 194

Search completed: October 27, 2002, 18:37:15  
Job time : 44.5337 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 14:35:03 ; Search time 123.369 Seconds  
(without alignments)  
6762.541 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486  
Sequence: 1 atgaacagagatgtttgca.....aagtcattcaacagaaaaa 486

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	486	100.0	486	22	AAF86246	OspA 17KD antigen
2	486	100.0	489	22	AAF79040	Piscirickettsia sa
3	295	60.7	483	22	AAF86247	DNA sequence of E
4	295	60.7	768	22	AAF86248	DNA sequence of C
5	73.2	15.1	110	22	AAF86252	PCR primer #2 used
6	71.6	14.7	118	22	AAF86254	PCR primer #4 used
7	69	14.2	102	22	AAF86255	PCR primer #5 used
8	62	12.8	94	22	AAF86253	PCR primer #3 used
9	56.2	11.6	454	20	AAZ06832	Rickettsia 17 KD o

C	10	46.8	9.6	110	22	AAF86256	PCR primer #6 used
C	11	44	9.1	309	23	ABL13399	Drosophila melanog
C	12	44	9.1	2309	23	ABL13298	Drosophila melanog
C	13	42.6	8.8	790	11	AAQ03870	H. influenzae DNA f
C	14	42.6	8.8	850	9	AAAB0227	Sequence of Haemop
C	15	40.4	8.3	1509	23	ABL20279	Drosophila melanog
C	16	40.4	8.3	5162	23	ABL20278	Drosophila melanog
C	17	39.8	8.2	2226	23	AA578586	DNA encoding novel
C	18	39.8	8.2	2292	21	AA578583	Arabidopsis thaliana
C	19	39.6	8.1	522	22	ABA49593	Human breast cell
C	20	39.6	8.1	522	22	ABA49593	Human foetal liver
C	21	39.6	8.1	522	22	ABA49593	Human foetal liver
C	22	39.6	8.1	522	22	ABA49593	Human brain expres
C	23	39.6	8.1	522	22	AAK41670	Human bone marrow
C	24	39.6	8.1	522	22	AAI22424	Probe #1257 for g
C	25	39.6	8.1	522	22	AAI47715	Probe #16401 used
C	26	39.6	8.1	522	22	AAI08110	Probe #8101 used t
C	27	39.4	8.1	1395	23	AA589524	DNA encoding novel
C	28	39	8.0	350	22	ABA72229	Human foetal liver
C	29	39	8.0	350	22	ABA38104	Probe #16570 for g
C	30	39	8.0	350	22	AAK20651	Human brain expres
C	31	39	8.0	350	22	AAK46795	Human bone marrow
C	32	39	8.0	350	22	AAI25744	Probe #15677 for g
C	33	39	8.0	350	22	AAI52634	Probe #21320 used
C	34	39	8.0	600	22	ABA59687	Human foetal liver
C	35	39	8.0	600	22	ABA28224	Probe #6690 for ge
C	36	39	8.0	600	22	AAK07956	Human brain expres
C	37	39	8.0	600	22	AAK33826	Human bone marrow
C	38	39	8.0	600	22	AAI16547	Probe #6480 for ge
C	39	39	8.0	600	22	AAI39552	Probe #8238 used t
C	40	39	8.0	801	23	AA567681	DNA encoding novel
C	41	39	8.0	8045	21	AA65171	Neurospora crassa
C	42	38.8	8.0	887	22	ABA51240	Human breast cell
C	43	38.8	8.0	887	22	ABA69248	Human foetal liver
C	44	38.8	8.0	887	22	ABA63164	Probe #14630 for g
C	45	38.8	8.0	887	22	AAK17542	Human brain expres

#### ALIGNMENTS

RESULT 1	AAF86246	standard; DNA: 486 BP.
ID	AAF86246	
XX	AAF86246:	
AC	11-JUL-2001	(first entry)
XX		
DT		
XX		
DE	OspA 17KD antigen gene.	
XX		
KW	Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;	
KW	vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;	
KW	SRS; ds.	
XX		
OS	Piscirickettsia salmonis.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..486
FT		/*tag= a
FT		/partial
FT		/product= "OspA"
FT		/note= "genus specific 17kDa antigen, the sequence does not include a stop codon"
XX		
CA281913-A1.		
PN		
XX		
PD	17-MAR-2001.	
XX		
PF	17-SEP-1999.	99CA-2281913.
XX		
PR	17-SEP-1999.	99CA-2281913.
XX		



OY 301 GATACAGCAATAGTATAGTGTGAGCCAGTCGACTTACACAGGCTTACATTAAGCAA 360  
 DB 301 GATACAGCAATAGTATAGTGTGAGCCAGTCGACTTACACAGGCTTACATTAAGCAA 360  
 OY 361 GAGCGTCGCCAGCAATATTGTGAGAAATTTACAGCAAAAGCGCATGATTCGAGGCGAGAA 420  
 DB 361 GAGCGTCGCCAGCAATATTGTGAGAAATTTACAGCAAAAGCGCATGATTCGAGGCGAGAA 420  
 OY 421 CAAAGCATTTAGCGCACTGCATGCCGCAACCGGATGCTGTTGGCAAGTCATTTACACA 480  
 DB 421 CAAAGCATTTAGCGCACTGCATGCCGCAACCGGATGCTGTTGGCAAGTCATTTACACA 480  
 OY 481 GAAAAA 486  
 DB 481 GAAAAA 486  
 RESULT 3  
 AAF86247  
 ID AAF86247 standard; DNA: 483 BP.  
 AC AAF86247;  
 XX  
 DT 11-JUL-2001 (first entry)  
 XX  
 DE DNA sequence of E. coli optimised ospa gene 17E2.  
 XX  
 KM Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KM vaccine; OSPA: salmonid rickettsial septicaemia; rickettsial disease;  
 KM SRS; 17E2; ds.  
 XX  
 OS Piscirickettsia salmonis.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..483  
 FT /tag- a  
 FT /product- "ospa"  
 FT /note- "Genus specific 17kDa antigen, the sequence does  
 not include a stop codon"  
 XX  
 PN CA2281913-A1.  
 XX  
 PD 17-MAR-2001.  
 XX  
 PF 17-SEP-1999; 99CA-2281913.  
 XX  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 PA (KAWW/) KAY W W.  
 PA (BURI/) BURIAN J.  
 PA (KUZV/) KUZV M A.  
 XX  
 PI Kay WW, Burian J, Kuzv MA;  
 XX  
 DR MPI: 2001-316844/34.  
 DR P-PSDB; AAB81127.  
 XX  
 PT Method for protecting polkiothermic fish against salmonid rickettsial  
 PT septicaemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the ospa protein of Piscirickettsia salmonis  
 XX  
 PS Example 3; Fig 4C; 35pp; English.  
 XX  
 CC This invention relates to a method for the protection against infection  
 CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of  
 CC OSPA in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkiothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents DNA which has been optimised for expression in  
 CC Escherichia coli to encode the P. salmonis ospa protein. An ospa protein  
 CC with an N-terminal fusion partner is used in a vaccine to create an  
 CC anti-ospa antibody response.  
 XX  
 SQ Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;  
 Query Match 60.7%; Score 295; DB 22; Length 483;  
 Best Local Similarity 76.0%; Pred. No. 2,3e-79;  
 Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
 OY 8 GAGGATGTTTGAAGGAGTAGTCTAATTTATTCAGTGTCTTTTATGTTGGTGGCC 67  
 DB 5 GTGTGTCCTGACAGGCGAGCTCTGTATCATTTATCTGTTTCTGTGGTGGTGGCC 64  
 OY 68 AGAATTTTACTGCTCAAGAAAGTCGAGCTGCGAGCTGCGGCTGTTGGCTGCTG 127  
 DB 65 AGAATTTTCAAGCCGCCAGGAAGTTGGCGCGCCACCGGTGGGTTGGCGGTGGCC 124  
 OY 128 GCCAGCTGTTTGTAAAGTAGTGTGAGTGCATGCAATGCCATTTGCTGTTTGG 187  
 DB 125 GCCAGCTGTTTGTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 184  
 OY 188 GTGATTTAATTGGTCTTAAATGCGTCAATGATGATGATGATGATGATGATGATG 247  
 DB 185 GCGGTCTGATTTGCTCTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 244  
 OY 248 ACCAGATTTTGAAGAAAGTAAAGCAGGCGCAAGTGCACCTTGGCTGATTCAGATAC 307  
 DB 245 ACCAGTCTGGAAGAAAGTAAAGCAGGCGCAAGTGCACCTTGGCTGATTCAGATAC 304  
 OY 308 GCAATAGTTTATGTTTGAAGCAGTGTGCTTACCAAGCTTTCATTAAGCAAGAGCTG 367  
 DB 305 GTAAACAGCTACTGCTGGAAGCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 364  
 OY 368 GCCAGCAATATTGTCGGAATTTTACAGCAAAAGCGATGATGATGATGATGATGATG 427  
 DB 365 GTACGCACTACTGCTGCGGAATTTTACAGCAAAAGCGATGATGATGATGATGATG 424  
 OY 428 TTTACGCGCACTGCATGCCGCAACCGGATGCTGTTGGCAAGTCATTTCAACAGAAAA 486  
 DB 425 TCTACGCGCACGGGTGCTGCTGACCGCGATGCGCTGCGAGTGATGATGATGATGATG 483  
 RESULT 4  
 AAF86248  
 ID AAF86248 standard; DNA: 768 BP.  
 AC AAF86248;  
 XX  
 DT 11-JUL-2001 (first entry)  
 XX  
 DE DNA sequence of c17e2 ospa construct with N-terminal fusion partner.  
 XX  
 KM Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KM vaccine; OSPA: salmonid rickettsial septicaemia; rickettsial disease;  
 KM SRS; 17E2; fusion construct; ds.  
 XX  
 OS Piscirickettsia salmonis.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..768  
 FT /tag- a  
 FT /product- "c17e2 ospa with N-terminal fusion partner"  
 FT /note- "No stop codon is given"  
 FT misc\_feature 1..285  
 FT /tag- b  
 FT /note- "DNA encoding undefined N-terminal fusion partner"  
 FT misc\_feature 286..768  
 FT /tag- c

/note="Optimised OspA construct c17E2"

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAYW/) KAY W W.  
(BURI/) BURIAN J.  
(KUZV/) KUZV M A.

Kay WW, Burian J, Kuzv MA;

WPI; 2001-316844/34.

P-PSDB; AAB81128.

Method for protecting poikilothermic fish against salmonid rickettsial  
vaccine containing the OspA protein of Piscirickettsia salmonis

Example 4; Fig 5; 35pp; English.

This invention relates to a method for the protection against infection  
of a poikilothermic fish by the bacterial pathogen, Piscirickettsia  
salmonis. The method comprises administering an immunogenic amount of a  
P. salmonis specific antigen termed OspA, or an immunogenic amount of a  
P. salmonis in the form of a vaccine. The method is used for protecting animals,  
particularly poikilothermic fish, against the bacterial pathogen  
P. salmonis. The method is also useful for protecting against salmonid  
ricketsial septicæmia (SRS) and other rickettsial diseases. The present  
invention represents P. salmonis OspA DNA termed C17E2 optimised for  
N-terminal fusion partner. The protein encoded by this fusion construct  
is used in a vaccine to create an anti-OspA antibody response.

Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;

Query Match 60.7%; Score 295; DB 22; Length 768;  
Best Local Similarity 76.0%; Pred. No. 2.8e-79;  
Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```

OY 8 GAGACTTTTGGCAGAGTAGTACTAATTAATATCACTGTCTTTTAACTGGCTGCTGCC 67
DB 290 GTGCTTCCCGCAGGCGCAGCTCTGATCATTAATCTGTTTCCCTGGTGGTGGCC 349
OY 68 AGAATTAGTGTCAAGAGTCCGAGCTGCGACTGGGGCTGTTGGCGGCTTGGCTG 127
DB 350 AGAATTAGTGTCAAGAGTCCGAGCTGCGACTGGGGCTGTTGGCGGCTTGGCTG 127
OY 128 GCGAGCTGTTGGTAAGAGTGTGCTGAGTGTGCAATGGCCATTTGGTGTGTTGG 409
DB 410 GCGAGCTGTTGGTAAGAGTGTGCTGAGTGTGCAATGGCCATTTGGTGTGTTGG 187
OY 188 GTGATTAATTGGTCTTAATCGTCAATGATGATGATGATGATGATGATGATGATG 469
DB 470 GCGGTCTGATTTGGCTTAATCGTCAATGATGATGATGATGATGATGATGATGATG 247
OY 248 ACCAGAGTTGGAAGAGTAAAGCAGGCAATGACAGTGGCGTAAATCCAGTACAG 307
DB 530 ACCAGAGTTGGAAGAGTAAAGCAGGCAATGACAGTGGCGTAAATCCAGTACAG 307
OY 308 GCAATGTTAGTGTGAGCGGCTGCTACTTACAGCGTTACATTAAGCAAGAGGCGC 589
DB 590 GCAATGTTAGTGTGAGCGGCTGCTACTTACAGCGTTCATTAAGCAAGAGGCGC 367
OY 368 GCGAGCAATTTTGTGAGATTTTCAAGAAAGGCGATGTTGAGGCGAGAGCAAGAGA 649
DB 650 GTAGCAATTTTGTGAGATTTTCAAGAAAGGCGATGTTGAGGCGAGAGCAAGAGA 427
OY 428 TTACGCACTGATGCGCGCAACCGGATGCTGTTGCAAGTCAATTTCAACGAAAA 709
DB 428 TTACGCACTGATGCGCGCAACCGGATGCTGTTGCAAGTCAATTTCAACGAAAA 486

```

Db 710 TCTAGGCAACCGCGTCCCTCAGCGGATGCGCGTGGAGGTGATTAACACGAAAA 768

RESULT 5  
ID AAF86252  
AAAF86252 standard; DNA; 110 BP.

AC AAF86252;  
DE 11-JUL-2001 (first entry)

DE PCR primer #2 used in cloning an optimisation of OspA gene.

KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
KW vaccine; OspA; salmonid rickettsial septicæmia; rickettsial disease;  
KW SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.  
PA (BURI/) BURIAN J.  
PA (KUZV/) KUZV M A.

PI Kay WW, Burian J, Kuzv MA;

WPI; 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial  
vaccine containing the OspA protein of Piscirickettsia salmonis

Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection  
of a poikilothermic fish by the bacterial pathogen, Piscirickettsia  
salmonis. The method comprises administering an immunogenic amount of a  
P. salmonis specific antigen termed OspA, or an immunogenic amount of a  
P. salmonis in the form of a vaccine. The method is used for protecting animals,  
particularly poikilothermic fish, against the bacterial pathogen  
P. salmonis. The method is also useful for protecting against salmonid  
ricketsial septicæmia (SRS) and other rickettsial diseases. The present  
invention represents a PCR primer used in the cloning and optimisation of  
the P. salmonis OspA gene. The OspA gene is used in the method of the

Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

Query Match 15.1%; Score 73.2; DB 22; Length 110;  
Best Local Similarity 79.1%; Pred. No. 1.9e-12;  
Matches 87; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```

OY 56 TTGGCTGTGCGCCAGAACTTGTGCTCAAGAGTGGAGCTGGAGTGGCTGTTGG 115
DB 1 TGGGTGGCGCCAGAACTTGTGCTCAAGAGTGGAGTGGAGTGGCTGTTGG 115
OY 116 GCGGTGTGTCGCGCCAGCTGTTGTAAGTGTGAGTGTGATGATG 165
DB 61 GCGGTGTGTCGCGCCAGCTGTTGTAAGTGTGAGTGTGATGATG 110

```

RESULT 6  
ID AAF86254/c  
AAAF86254 standard; DNA; 118 BP.



AC	AAf86254;
XX	
DT	11-JUL-2001 (first entry)
DE	PCR primer #4 used in cloning an optimisation of OspA gene.
XX	
KW	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM	vaccine; Osipa; salmonid rickettsial septicaemia; rickettsial disease;
XX	SRS; 17E2; fusion construct; PCR primer; ss.
OS	Piscirickettsia salmonis.
PN	CA2281913-Al.
XX	
PD	17-MAR-2001.
PF	
XX	17-SEP-1999; 99CA-2281913.
PR	
XX	17-SEP-1999; 99CA-2281913.
PA	(KAYW/) KAY W W.
XX	(BURI/) BURIAN J.
XX	(KUZY/) KUZYK M A.
PI	Kay WW, Burian J, Kuzyk MA;
DR	WPI: 2001-316844/34.
XX	
PT	Method for protecting poikilothermic fish against salmonid rickettsial
PT	septicaemia and other rickettsial diseases comprises administering a
PT	vaccine containing the OspA protein of Piscirickettsia salmonis -
XX	
PS	Example 3; Fig 4B; 35pp; English.
XX	
CC	This invention relates to a method for the protection against infection
CC	of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC	salmonis. The method comprises administering an immunogenic amount of a
CC	P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC	OspA in the form of a vaccine. The method is used for protecting animals,
CC	particularly poikilothermic fish, against the bacterial pathogen
CC	P. salmonis. The method is also useful for protecting against salmonid
CC	rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC	sequence represents a PCR primer used in the cloning and optimisation of
CC	the P. salmonis OspA gene. The OspA gene is used in the method of the
CC	invention.
SO	
XX	Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
XX	
Query Match	14.7%; Score 71.6; DB 22; Length 118;
Best Local Similarity	75.4%; Pred. No. 5.9e-12;
Matches 89; Conservative 0; Mismatches 29; Indels 0; Gaps 0	
OY	211 GGTCATCGATGATCAGCAGAGATAAATAAAGCTAAACCACAGTTGGAAAAAGTAAAA 270
DB	118 GGTCAAGCATGAGCACAGAGATTAACAACACTGAACACAGCTCTGGAAAAAGTGAAA 59
OY	271 GCAGGCGAAGTGACACGTTGGCGTATATCCAGATTACAGGCACATATGTTAGTGTGGAGC 328
DB	58 GCCGCCACAGGTACTCGTTGGCGTATCCGACACCGGTAACAGCATCTGTGGAAC 1
RESULT 7	
ID	AAF86255/C
XX	AAF86255 standard; DNA; 102 BP.
AC	AAf86255;
XX	
DT	11-JUL-2001 (first entry)
DE	PCR primer #5 used in cloning an optimisation of OspA gene.
XX	
KW	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM	vaccine; Osipa; salmonid rickettsial septicaemia; rickettsial disease;

KW	SRS: 17E2: fusion construct: PCR primer: ss.
XX	
OS	<i>Piscirickettsia salmonis</i> .
XX	
PN	CA2281913-A1.
XX	
PD	17-MAR-2001.
XX	
PF	17-SEP-1999; 99CA-2281913.
XX	
PR	17-SEP-1999; 99CA-2281913.
XX	
PA	(KAYW/) KAY W W.
XX	
PA	(BURJ/) BURIAN J.
XX	
PI	(KUZV/) KUZRYK M A.
XX	
PI	Kay WW, Burian J, Kuzzyk MA;
XX	
DR	WPI: 2001-316844/34.
XX	
PT	Method for protecting polikilothermic fish against salmonid rickettsial
XX	
PT	septicemia and other rickettsial diseases comprises administering a
XX	
PT	vaccine containing the Ospa protein of <i>Piscirickettsia salmonis</i>
XX	
PS	Example 3; Fig 4B; 35pp: English.
XX	
CC	This invention relates to a method for the protection against infection
XX	
CC	of a polikilothermic fish by the bacterial pathogen, <i>Piscirickettsia</i>
XX	
CC	salmonis. The method comprises administering an immunogenic amount of a
XX	
CC	P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
XX	
CC	Ospa in the form of a vaccine. The method is used for protecting animals,
XX	
CC	particularly polikilothermic fish, against the bacterial pathogen
XX	
CC	P. salmonis. The method is also useful for protecting against salmonid
XX	
CC	rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX	
CC	sequence represents a PCR primer used in the cloning and optimisation of
XX	
CC	the P. salmonis Ospa gene. The Ospa gene is used in the method of the
XX	
CC	Invention.
XX	
SQ	Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;
XX	
Query Match	14.2%; Score 69; DB 22; Length 102;
Best Local Similarity	80.2%; Pred. No. 3.4e-11;
Matches 81; Conservative	0; Mismatches 20; Indels 0; Gaps 0.
OY	337 ACTTACACGCGTTACAAATTAAGCAAGAGCGTCGCGACGATATTGTGAGAAATTTCAGCAA 396
DB	102 ACGTACACGCGTTACAAACAAACAGGAGCGCGTCGACGACTACTGCCGCGAATTTTCAGCAG 43
OY	397 AAGGCGATGATTGCAGGCGAGCAAGCAAGATTTACGCGAC 437
DB	42 AAAGCGATGATCGCAGGTCGAAACAGGAATCTACGCGAC 2
RESULT 8	
AAAF86253	
ID	AAAF86253 standard; DNA: 94 BP.
XX	
XX	AAF86253;
XX	
XX	11-JUL-2001 (first entry)
DE	PCR primer #3 used in cloning an optimisation of Ospa gene.
XX	
KW	Polikilothermic fish: <i>Piscirickettsia salmonis</i> ; rickettsial pathogen;
XX	
KW	vaccine: Ospa, salmonid rickettsial septicemia; rickettsial disease;
XX	
KW	SRS: 17E2: fusion construct: PCR primer: ss.
XX	
OS	<i>Piscirickettsia salmonis</i> .
XX	
XX	CA2281913-A1.
PN	
XX	
PD	17-MAR-2001.
XX	

of a poikilothermic fish by the method for the protection against infection salmonis. The method comprises administering a bactericidal pathogen, *Piscirickettsia salmonis* specific antigen termed OspA, or an immunogenic amount of OspA in the form of a vaccine, against an immunogenic amount of *P. salmonis*. The method is also useful for protecting against rickettsial septicaemia (SRS) and other rickettsial diseases. The present invention represents a PCR primer, used in the cloning and optimisation of the *P. salmonis* OspA gene. The OspA gene is used in the method of the

Query Match	Best Local Similarity	Score 62	DB 22	Length 94
Matches 74, Conservative	12.8%;	78.7%;	Pred. No. 4.3e-09;	
			Mismatches 20, Indels 0, Gaps 0;	
142 AAAGGTGCTGCTGACCTTGCAATGGCCATGCTGGTGTGCTGCTGTTTGGTGGCATTAATGCT				201
1 AAAAGCTCTGGTCTGGTGGCATGCGCATGGGGGGGCGGTGTGGGGCGTCTGATTGTC				60
202 TCTAAATATGGTCAATCGATCGATGATGACGAGATA				235
61 TCTAAATATGGTCAAGACCATGAGCAGCAGAGATA				94

LT 9  
 66832  
 AA206832 standard; DNA; 454 BP.  
 AA206832;  
 09-NOV-1999 (first entry)  
 Rickettsia 17 kD outer membrane protein DNA.  
 Bacteria; intracellular; parasite; Outer membrane protein; OMP; vaccine  
 diagnosis; immune response; antigen; peptide; tick; spotted fever; ds.  
 Rickettsia helvetica.  
 ey  
 DSS Location/Qualifiers  
 1..454  
 /\*tag= a  
 /partial  
 /product= "17 kD outer membrane protein"  
 /note= "Stop codon not given in specification"  
 277..360  
 /\*tag= b  
 /product= "28 amino acid antigenic peptide from  
 17 kD OMP"

CC Rickettsia represents a major portion of the coding sequence for  
CC Rickettsia helvetica 17 kD outer membrane protein (OMP) and includes the  
CC sequence encoding a 28 amino acid antigenic protein (OMP) and includes the  
CC PCR primers 17up (AAZ068933) and 17LOW (AAZ068942) Rickettsia using  
CC a member of the spotted fever group (SFG) of the Rickettsia genus of  
CC Gram negative bacteria. SFG Rickettsiae comprise about 20 different, but  
CC closely related, species and are transmitted to humans via ticks.  
CC Rickettsiae are obligate intracellular parasites in both ticks and  
CC humans. The Rickettsiae found in Ixodes ricinus ticks are pathogenic for  
CC humans, causing high fever, headache and rash, plus other,  
CC pathogen-specific symptoms. However, the symptoms of rickettsiosis are  
CC similar to some other bacterial and viral infections. The immunogenic  
CC properties of Rickettsia bacteria are associated with structures exposed  
CC on the surface of the organism. Lipopolysaccharides (LPS) have been used  
CC as the basis for a serological assay. The well-felix reaction, but is  
CC very non-specific. use of the 28 amino acid antigenic protein, but is  
CC 17 kD OMP provides a more accurate diagnosis of rickettsial infection.  
CC Rickettsia helvetica or very closely related species in body fluid  
CC samples. It can also be used in vaccines to protect against, or treat,  
CC rickettsiosis.  
XX Sequence 454 bp. 129

Query Match	Best Local Similarity	11.6%	Score 56.2	DB 20	Length 454
Matches 182	Conservative	0	Mismatches 188	Indels 3	Gaps
Qy	82 CAAGAATCGAGCTGGCACTGGGGCTGTGTTGGCGCTGTCTGCGCAGCTGTTGGT	11	11	11	11
Db	85 CAAGTACAGGAACACTCTTGGCCGCTCCGGCGGTCACTTCTGTTGTTCAATTGGT	11	11	11	11
Qy	142 AAAGTGTGTTGAGTTTGCATATGGCATGGCCATTGGTGTCGTTGGTGGTCAATTTGGT	11	11	11	11
Db	145 AAAGCTAAAGGCAACTGTGTGGAGTGTAGTGTGATGCTACTTATGGAGCAGTTCTTGGC	11	11	11	11
Qy	202 TCTAAATCGCTCAATCGATGATCGACAGAGATTAATTAAGCTAAACCGAGTTGGCA	11	11	11	11
Db	205 GGCGCAATCTGTTGACGATATGATGACGAGTAAACACTTGGACGTTTCCGCAC	11	11	11	11
Qy	262 AAGGTAAAGCAGGCGCAAGTGACACTTGGCGTATCCAGATACAGCAATAGTTATAGT	11	11	11	11
Db	265 AAGCTTTTAGAAGC---AGTCTTACCGGTGTGTAACTGAGTGGCGTAAATCCGATAC	11	11	11	11
Qy	322 GTTGAACAGTGGCTACTTACCAAGGTTATCAATATACAGAGCGTCCGACCAATATTGT	11	11	11	11
Db	322 GGCATTTAGCGTTACGTAAACCTTAATMAAATTATAGAAATAGACATGGTCAATATTGC	11	11	11	11
Qy	382 CGAGAAATTCACCAAAAGCGATGTTGCAGGGCAGAGAACAGATTTCCAGGCATGCA	11	11	11	11
Db	382 CGTGAATACACTCAAAACAGTTGTTAATAGCGGAAAAACAACAAAAGCATACGTTATGCA	11	11	11	11
Qy	442 TGGCGCAACCGG 454	11	11	11	11



[illegible]

```

XX      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signaling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins
CC      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pcl_sequences.
SQ
SQ      Sequence 1509 BP; 495 A; 446 C; 277 G; 291 T; 0 other;
      Query Match      8.3%; Score 40.4; DB 23; Length 1509;
      Best Local Similarity 49.1%; Pred. No. 0.053;
      Matches 107; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY      8 GAGGATGTTTGCAGAGTAGTACTAAATTATTATCAGTGTGTTTTTAACTGGCTGTGCC 67
DB      985 GTGGTCTGTGTGGTGTCTGTGTGGAGGTTCTGTGGTTCGTGCGTGCCTCCGTCG 926
QY      68 AGACTTTAGTGTCTCAAGACGCGAGCTGCGACTGGGGCTGTCTTGGCGGCTGTGCTG 127
DB      925 GTGGCTTTGTGTGGTGTCTTAACTTGGAGCAATTGTGTGAGATTAATGTGTGGAGCAATTGTGG 866
QY      128 GCCACCTGTTTGTAAAGGTACTGTGTCGAGTTGCAATGGCCATTGTGTGTGCTGTTTTGG 187
DB      865 GAGGAATTGTGTAAGCAATTGTTGGAGGATTTGTGATGCTTCCTGTGGACTAGTCG 806
QY      188 GTGGATTAAATTGGTCTTAAATCGGTCAATCGATCGAT 225
DB      805 GGGGAAGTGTGGGTGATAGTAGGTGATGTGTGTGT 768

Search completed: October 27, 2002, 16:40:39
Job time : 127.389 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:07:23 ; Search time 941.223 Seconds  
(without alignments)  
10805.408 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486  
Sequence: 1 atgaacagagagatgttgca.....aagtcattcaacagaaaa 486

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pal:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	486	100.0	489	6	AX252413	AX252413 Sequence
2	486	100.0	4983	1	AF184152	AF184152 Piscirickettsia
3	79	16.3	448	1	RSU76907	RSU76907 Rickettsia
4	70.6	14.5	237523	1	RPRX04	AJ235273 Rickettsia
5	68.4	14.1	416	1	AF031534	AF031534 Rickettsia
6	65.8	13.5	535	1	RIRANT17KB	M28481 R. typhi 17K
7	65.8	13.5	537	1	RIR17KCA	D16515 Rickettsia
8	64.6	13.3	539	1	RIRANT17KA	M28480 R. conorii 1
9	64.6	13.3	10127	1	AE008675	AE008675 Rickettsia
10	64.2	13.2	620	1	RIRANT17KD	M28479 R. Rickettsia
11	64.2	13.2	620	1	RIRANT17K	M16486 R. Rickettsia
12	62.2	12.8	434	1	MR1269516	AJ269516 Male-Kill
13	62.2	12.8	434	1	MR1269517	AJ269517 Male-Kill
14	62.2	12.8	434	1	MR1269518	AJ269518 Male-Kill
15	59.4	12.2	546	1	AF195118	AF195118 Rickettsia
16	59	12.1	489	1	RROU1020	U11020 Rickettsia
17	58.2	12.0	415	1	AF027124	AF027124 Rickettsia
18	57.8	11.9	532	1	RIRANT17KC	M28482 R. prowazeki
19	57.6	11.9	411	1	AF060704	AF060704 Rickettsia
20	57.6	11.9	411	1	AF060706	AF060706 Rickettsia
21	57.4	11.8	489	1	RPU17008	U17008 Rickettsia
22	57.4	11.8	489	1	RAU11013	U11013 Rickettsia
23	56.2	11.6	454	1	AF181036	AF181036 Rickettsia
24	55.6	11.4	489	1	RM011017	U011017 Rickettsia
25	54.4	11.2	395	1	U04162	U04162 Rickettsia
26	54.2	11.2	492	1	RIRTRAPRO	M74042 R. australis
27	52.6	10.8	491	1	RIRGENSEO	M99391 Rickettsia
28	47.8	9.8	394	1	AF260571	AF260571 Rickettsia
29	47.8	9.8	7218	6	166494	166494 Sequence 14
30	47.4	9.8	63454	2	AC102004	AC102004 Mus muscu
31	46.4	9.5	111824	9	AL160412	AL160412 Human DNA
32	45.4	9.3	68650	9	AC027343	AC027343 Homo sapi
33	45.4	9.3	123670	2	AC103214	AC103214 Rattus no
34	45	9.3	156370	2	AL158162	AL158162 Homo sapi
35	45	9.3	166464	9	AL158212	AL158212 Human DNA
36	44.6	9.2	13123	1	AE003872	AE003872 Xylella f
37	44.6	9.2	46877	9	AC104133	AC104133 Homo sapi
38	44.4	9.1	184999	2	AL611934	AL611934 Mus muscu
39	44.2	9.1	474	3	PFU2952	U72952 Plasmodium
40	44.2	9.1	475	3	AF104714	AF104714 Plasmodiu
41	44.2	9.1	585	3	PFAMS21	L19047 Plasmodium
42	44.2	9.1	615	3	PFAMS21	L19045 Plasmodium
43	44	9.1	68418	2	AC018090	AC018090 Drosophila
44	44	9.1	153048	3	AC008362	AC008362 Drosophila
45	44	9.1	175302	2	AC020587	AC020587 Homo sapi

ALIGNMENTS

RESULT 1  
AX252413  
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 5 from Patent WO0168865.  
ACCESSION AX252413  
VERSION AX252413.1 GI:15985721  
KEYWORDS  
SOURCE  
ORGANISM  
Piscirickettsia salmonis.  
Piscirickettsia salmonis  
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;  
Piscirickettsia.  
REFERENCE  
1 (bases 1 to 489)  
Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and  
Burzio,L.  
TITLE Fish vaccine against piscirickettsia salmonis  
JOURNAL Patent: WO 0168865-A 5 20-SEP-2001;  
Aqua Health (Europe) Limited (GB)  
FEATURES  
source  
1. 489  
/organism="Piscirickettsia salmonis"  
/db\_xref="taxon:1238"  
BASE COUNT 139 a 79 c 144 g 127 t  
ORIGIN

CDS

0;

```
<1. .501
/gene="alr"
/codon start=1
```

gene  
complement(526. .1257)  
/gene="Bax"  
CDS  
complement(526. .1257)  
/gene="Bax"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="Bax"  
/protein\_id="AAC17002.1"  
/db\_xref="GI:1041348"  
/translation="MELSAVITTAVALAOPLEFSAPTAKTQITNADISQISIKQFEIIL  
LKLRVDITIPALITSAQAINESNMWGSRAVAGNNFGMCHIKLNNLHFPHDDFS  
GKRYVELLQITITTAIHTLINTNAYOALRDLRAHMRANQDVSAFKLAEGITVYSIK  
1373. .2744"  
/gene="radr"  
gene  
evidence=experimental  
1373. .2744

gene  
2834. .3322

ORGANISM	CDS	/evidence-experimental
<i>Piscirickettsia salmonis</i>	2834	3220
<i>Rickettsia salmonis</i>	2834	3220

```

... 17 kDa antigen"
/codon_start=1
/transl_table=11

```

JOURNAL of Fish Diseases 1990, 13, 101-105

**Authors** Kuzzyk, M.A. and Kay, W.W.  
**Title** Direct Submission

**FEATURES**  
**source**  
 Location/Qualifiers  
 1. 4983  
 Tech Building, PO Box 3055, University of  
 Victoria, BC V8W 3P6, Canada

```
gene /clone="pBI2"  
      <1. .501  
      /gene="a1r"
```

```
repeat_region      3560..3846      insertion sequence-like element
gene               complement(3842..4561)
CDS                /gene="trpA"
                   complement(3842..4561)
                   /gene="trpA"
                   /codon_start=1
```



```
repeat_region
BASE COUNT      1362 a      778 c      1265 g      1578 t
ORIGIN
Query Match      100.0%; Score 486; DB 1; Length 4983;
Best Local Similarity 100.0%; Pred. No. 6.3e-123;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/transl_table=1
/product="transposase"
/protein_id="AAG16999.1"
/db_xref="GI:10441345"
/translation="MDLJLISLFCVIDPFOELLIPONAILLEDYTKRKRPSPOMSTS
EIMIMTYEHRKSNRYNRYLHVHVKSNVYKFPNSVSYNRYEMLPSILPLCFETA
AOGATATGIVFDSTILRVCHKEKASQNRAPKGLAKSKSTMGWYGFKHIIVNDG
ELMAFKMSKATTDSDRVVLPKMAENLTKIIGDKYISQKLFDOYERGLDITIRKN
MKNKLVIMIKIKILKKRGIIA"

QY 1 ATGAACAGAGATGTTTTCAGAGTAGTCTAATTATTATTCAGTGTGTTTGTAGTGGC 60
Db 2834 ATGAACAGAGATGTTTTCAGAGTAGTCTAATTATTATTCAGTGTGTTTGTAGTGGC 2893

QY 61 TGTGCCCGAAGCTTTAGTCGCAAGAGTCGAGCTGCGACTGGGGCTGTTGGCGGT 120
Db 2894 TGTGCCCGAAGCTTTAGTCGCAAGAGTCGAGCTGCGACTGGGGCTGTTGGCGGT 2953

QY 121 GTTGCTGCGCAGCTGTTGGTAAAGGTAGTGTGAGTGTGCAATGCGCATGGTGCT 180
Db 2954 GTTGCTGCGCAGCTGTTGGTAAAGGTAGTGTGAGTGTGCAATGCGCATGGTGCT 3013

QY 181 GTTTGGGTGATTAAATGTTCTAAATCGTCATCGATGATGATCAGACAGATAAATA 240
Db 3014 GTTTGGGTGATTAAATGTTCTAAATCGTCATCGATGATGATCAGACAGATAAATA 3073

QY 241 AAGCTAAACCGAGATTTGGAAAAAGGTAAACAGCGGCAAGTGCACGTTGGCGTATCA 300
Db 3074 AAGCTAAACCGAGATTTGGAAAAAGGTAAACAGCGGCAAGTGCACGTTGGCGTATCA 3133

QY 301 GATACAGCAGATAGTATAGTGTGAGCGGCTACTTCCAGCCGTACCAATTAAGCAA 360
Db 3134 GATACAGCAGATAGTATAGTGTGAGCGGCTACTTCCAGCCGTACCAATTAAGCAA 3193

QY 361 GAGGCTGCCAGCAATATTTGTGAGAAATTTGACAAAAGGCGATGATTCAGGCGAGAG 420
Db 3194 GAGGCTGCCAGCAATATTTGTGAGAAATTTGACAAAAGGCGATGATTCAGGCGAGAG 3253

QY 421 CAAGAGATTTACGGCAGTCGATGCCGCGAACCGGATGCTGTGGCAAGTCAATTTCAACA 480
Db 3254 CAAGAGATTTACGGCAGTCGATGCCGCGAACCGGATGCTGTGGCAAGTCAATTTCAACA 3313

QY 481 GAAAAA 486
Db 3314 GAAAAA 3319

RESULT 3
RSU76907      448 bp      DNA      linear      BCT 18-FEB-1998
LOCUS      Rickettsia sp. 17kha common-antigen gene, partial cds.
DEFINITION
VERSION      U76907.1      GI:2894794
KEYWORDS
SOURCE
ORGANISM      Rickettsia sp.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia.
REFERENCE
AUTHORS      Davis,M.J., Ying,Z., Brunner,B.R., Pantoja,A. and Ferwerda,F.H.
TITLE      Rickettsial relative associated with papaya bunchy top disease
JOURNAL      Curr. Microbiol. 36 (2), 80-84 (1998)
MEDLINE      98087556
REFERENCE
AUTHORS      Ying,Z. and Davis,M.J.
TITLE      Direct Submission
```

```
JOURNAL      Submitted (01-NOV-1996) Tropical Research and Education Center,
University of Florida, 18905 SW 280 Street, Homestead, FL 33031,
USA
FEATURES
Source      Location/Qualifiers
1. .448
/organism="Rickettsia sp."
/db_xref="taxon:789"
/note="Puerto Rico"
<1. .>448
/note="putative"
/codon_start=3
/transl_table=1
/product="17 kDa common-antigen"
/protein_id="AAG02809.1"
/db_xref="GI:2894795"
/translation="ALNASHMLEACQGGPGNNKQSGTLIGTIGLVGSGFGGCTGR
LAAGAGALIGAILGNOIGAGMDEDRILATISORALEAPSSSVWRNPDPNGVGT
VPSKAVKNNTGVCREYTOTVYVGGQOKAVGTACRQPDGQWV"

BASE COUNT      135 a      81 c      120 g      112 t
ORIGIN
Query Match      16.3%; Score 79; DB 1; Length 448;
Best Local Similarity 51.6%; Pred. No. 5e-11;
Matches 206; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 72 CTTAGTGTCAAGACAGTCGAGCTGCGACTGGGGCTGTTGGCGGTGCTGGCCA 131
Db 47 CATGAATACAAAGTAGTGTGCTGCAATGTCGGAACCTTTAGAGAGACTGTTGGCTTC 106

QY 132 GCTGTTGGTAAAGTAGTGTGCTGCAATGTCGGAACCTTTGGTGGGTGG 191
Db 107 ACAATTTGGTGTGTACGGGACGACTGTCGGGTAGACAGACAGTGTACTTGGTGC 166

QY 192 ATTATTTGTTCTAAATCGTCAATCGATGATCAGCAGATTAATTAAGCTTAACCA 251
Db 167 AATCCTTGGCAACCAAAATGTTGTCAGATGATGAACAAATGTAAGAACTTGCAGAGCT 226

QY 252 GAGTTTGGAAAGGTAAAGCAAGGCAAGTGAACGTTGGGTAAATTCAGATPACAGCAA 311
Db 227 TACTTCTCAAGACCTTTAGAG--CTGCACCAACCGGTGACGTGTCAATGGCCTTA 283

QY 312 TAGTTATAGTGTAGCCAGTGCCTACTTACCAGCGTTCAATTAAGCAAGAGCTCGCA 371
Db 284 TCCGATATATGTTATATGTTACTGTACACCAAGCAACCTTAATAAATAATPACGG 343

QY 372 GCAATATTGTGAGATTTACGCAAAAGCGATGTTGACAGGCAAGAACAGATTTA 431
Db 344 TCAATATTGTGTAATACACCAACAGTTGTAGTAGTGAAACAAACAAAGCTTA 403

QY 432 CGGCACTGCATCGCGGCAACGGGATGTCGTTGGCAAGT 470
Db 404 TGGCACTGCTGCGGCTCAACCTGATGACAAATGGCAAGT 442

RESULT 4
RPXX04
LOCUS      Rickettsia prowazekii strain Madrid E, complete genome; segment
DEFINITION
ACCESSION      AJ235273      AJ235269
VERSION      AJ235273.1      GI:3861237
KEYWORDS      complete genome.
SOURCE
ORGANISM      Rickettsia prowazekii.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.
REFERENCE
AUTHORS      Andersson,S.G., Zomorodipour,A., Andersson,J.O.,
Sicheritz-Ponten,T., Alsmark,U.C., Podewski,R.M., Naslund,A.K.,
Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
TITLE      The genome sequence of Rickettsia prowazekii and the origin of
mitochondria
JOURNAL      Nature 396 (6707), 133-140 (1998)
```

**MEDLINE** 99039499  
**REFERENCE** 2 (bases 1 to 237523)  
**AUTHORS** Andersson, S.G.E.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (11-NOV-1998) S.G.E. Andersson,  
 Siv.Andersson@molbio.uu.se, Dept. of Molecular  
 Biology, Uppsala, S-751 24, SWEDEN  
**FEATURES** Location/Qualifiers  
**source** 1 372

```

/strain="Madrid E"
/db_xref="taxon:782"
164. .1123
/gene="RP701"
164. .1123

```

```

/obj_xref=10="CA515137.1"
/obj_xref=61="3861238"
/translation="MAEPPIETLEGGKDKIDLRINNOALNKLTHTPLDQWKEHSIH
ALPQSELSKXWOLTEKDDHEFTNIDNNWVKTIPADOLIAONLNKTKTITTEFYKVA
LNIINHEHNGSSSTIKTKVOKLTKTKANDOSKEILTEWPIKOTVTTITDRIATVSE
EDQJQKLEIHEILKDNNAEAFENKQYVSOISROFKKDDKIDDKOAHNIDSNENQKVT
NDSKQIVQVHTITSYVQDKEKOSDGGFELTRLEPNTFKNYEAKO"
2074..3348

```

[illegible]

/protein\_id="CA15139.1"  
/db\_xref="GI:3861240"  
/translation="MSKIGNPFLITMTCLSIITLIIIPWLDIIRKRRNNICIKFE  
PWICTSGQVQVATITFYASLCSITAFQETLVAVITDFDEVNPLFHSSTLKIYRI  
SGMSHIEGMLMPLDIALVSCYHLEMLEMKRSITLISALQILFSSRIYPTSN  
PFHNEFVEPEIGIGPMPLDIALHPLPLTIGVGYAVPFTVTCALFEIIVTSL  
NMLPEFETGKTQICLSIDRTSITLIPFACGQILFTTTEISGSMWAAFEIILWGLSD  
EDPENISITPMLSSIMMHSHIITWTKTMOHWITITLITFELVITSTPLVSGI  
NSTHIAARSESEMYITLVLCIITIDGGLVYIIRLNCSTIEKTYITFPMWNSHCNT  
IKRKALLIGNFEFLSITLICALIATPYTITILNCSITLTIYVITFSTPLVSGI  
TIGLFTNLSISKIKHIMLSITITPYTITILNCSITLTIYVITFSTPLVSGI  
KTLFTRNSISLSSKISMLHGVALFASITFNSIILTMISFLMATHAVYLV  
LONIFAGKNGYQVQINPMLDKNHKNITLIDENKREYITLQKLSQSDSHSLAYLD  
YAVSLDGIITIAHKITFKMSFEIWIQVITTTAGVAILRNKSY"  
complement(7264..12195)  
(gene="88704..12195)

gelle  
CDS

misc\_feature

...: .13216  
/gene="RP705"

gene

9. **முனைய**

CDS

/protein\_id="CA15140.1"  
 /db\_xref="GI:3861241"  
 /translation="MAQPFNLKRIISAGLVASTATITVAGSGVAAQAAWVNRITN  
 INDDTITGLTITIAQAAEFNFVPAACKLTINIGQITITVQASNTINQAALIKVING  
 AAINANDISGLTSTITFAAPAEVLEFNIINPTQEPDLITLQASNTIVAGAGTLNIG  
 SKVGAATERNVAGKITITFAAPAEVLEFNIINPTQEPDLITLQASNTIVAGAGTLNIG  
 SVYDQNGATITVAGAGTLNIGSKVGAATERNVAGKITITFAAPAEVLEFNIINPTQEP  
 NRESSTENGLDQIIVVPTPKILKGNPTGVKNNNGTAITVFNNGALVASEVDN  
 IAVTINAIENEGAGVLESGHITIELNTAVACKLISQGAANAIVGTNDNGAGAGFI  
 LAAGSITQDGSALITIGGIGGVNALDHTILADASKILALDGANIIGANNGALHT  
 QANGSTIKLTNTQNTNIYVNDITLDTDKTGVDASKLTLNQNTLITIGASTIGYAT  
 LAQNTISSSTKITINAGDVAINLEITVETENNGSVQNLHNTYILITKITINAQCGITV  
 LNTQITLADDTNGLSGENPLSTHTETENNGSVQNLHNTYILITKITINAQCGITV  
 LHTPSSGTSIVSTGVCGQGHKLNNLLDNGTAVKFLNCGKGVANITITNDASV  
 NYTTHDYESADNTGTELEPNTDLITVTLNKGQAVKFEVLQVILISPGVITNEIGI  
 IVHQAISNTSEFNASICTSLFPLSGTPLNKGQAVKFEVLQVILISPGVITNEIGI  
 MINNGQITIDCKNTITLLETDGDSNIYANATITVSGTIRKGVGVDFMAPIVNEIGI  
 IYELGENSEPFLKQITVLETDGDSNIYANATITVSGTIRKGVGVDFMAPIVNEIGI  
 SVNGVANVPDSTITVLETDGDSNIYANATITVSGTIRKGVGVDFMAPIVNEIGI  
 DSAAGLIGTITVLETDGDSNIYANATITVSGTIRKGVGVDFMAPIVNEIGI  
 GGTATSTLWASVQNTIDETVYNTLINSNTILGGGTLNGLNINISITDLTPAASVFGN  
 AGVSONSLGKPNPAAVAGSNIPYKVELIDDSMDVVIKRDNIJVAATVVAQNSAI  
 ODLISNRLGTSLESTNPAANMMLLAKDPSDAITFGALITDYSAAVTVAQNSAI  
 IAGTKAKTGFQVAVGLTILASDLNIMIGALITGKITDIHODYKKGDDTINGLISB  
 SOQLKRTFQVQNAITFLAKKVKRSQRPFTSNCKMSKOIAGENDNMTFGCNILB  
 DYMAPVAVITPAGASTYLSKSNENYKRTGTGVANKRKNSKSDRVDLVAKAYAST  
 VNTIDITVPEIHSFVYHVKNTKSNOSMLDQITAFITSDPDPAKTASYNIGLSMT  
 KSDMLKMGYGGDFENASAKYTAHOGTILKVRNF"  
 complement(12953..13216)  
 /gene="RP705"  
 /note="Pseudogene that lacks ATG or GTG initiation codon.  
 Homology to guanine-3',5'-bis(diphosphate)  
 3-pyrophosphohydrolase)  
 complement(12953..13216)  
 /gene="RP705"  
 complement(13563..14507)  
 /gene="RP706"  
 complement(13563..14507)  
 /gene="RP706"  
 complement(13563..14507)  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA15141.1"  
 /db\_xref="GI:3861242"  
 /translation="MKIKOTAPYITISPELTDAERLEFEENHPICITILFRNIRK  
 NERGEDOKAALIKLILAKDPELVGNTITIALIDEGGVNRLIAPFYAPAAQTFTEL  
 TIKYNSVSTIAKLEFVGNINDFAVNDLHDADNIISDRFGKEPPIVAFTELO  
 GLOEENVACIKIHPIGHRNITVSHLELITPSLTKLETDIDPVFVLEPLSLA  
 MNAVHIVALEDBNDPATLSQIVITAKSNIGKGLIISDRIENQALISGSMADITGAT  
 DAGVAVILECGEENRKAISEVLSIDKPSDLIFS"  
 complement(14552..15025)  
 /gene="RP707"  
 complement(14552..15025)  
 /gene="RP707"  
 complement(14552..15025)  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA15142.1"  
 /db\_xref="GI:3861243"  
 /translation="MINNKLKKYVASEVIRILNALHKLRLVETKIPGLSHYKKN  
 KRYTVAHDIDLOKFLFSPREPELVKGRHSDDYLANSNHCAKSDPSRQNEQG  
 KNECITLHNVQORODSNAVYSLKMDNTYSRIDILLTNPNHLSIQIKKILTILS"  
 complement(15220..15507)  
 /gene="RP708"  
 complement(15220..15507)  
 /gene="RP708"  
 complement(15220..15507)  
 /codon\_start=1  
 /transl\_table=11  
 /product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (hima)"  
 /protein\_id="CA15143.1"

```

/db_xref="GI:3861244"
/translation="MTITKNKISIMLSKLGFSNN/CEAIVNTVFSNILEIKAVOKLT
LNKSGFEVKOKTTPRGINFHTKSPVMAISKNNLRTTPESEKALKALINKSML"
18403..18701
/gene="RP709"
18405..18701
/gene="RP709"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15144.1"
/db_xref="GI:3861245"
/translation="MAVGRYIEDHTPEOLHKLCDNLVKYLEETKAITHEKAEDTEV
VIEVSLDINSDDASIIKKKAITVSDSIANSNEDANNTKSHODVLRCTYQ"
complement(19685..20789)
/gene="RP710"
complement(19685..20789)
/gene="RP710"
/feature="Pseudogene with two authentic frameshifts and four
authentic in-frame termination codons. Homology to
transposase."
complement(21112..21608)
/gene="RP711"
complement(21112..21608)
/gene="RP711"
/feature="Pseudogene with one authentic frameshift. Homology
to resolvase."
complement(22602..23066)
/gene="RP712"
complement(22602..23066)
/gene="RP712"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA15145.1"
/db_xref="GI:3861246"
/translation="MSIERKEFMASHKQVKEIGTGNLNGKHDLFKAVAYTSDIEA

Query Match      14.5%: Score 70.6; DB 1; Length 237523;
Best Local Similarity 50.6%; Pred. No. 2.3e-08;
Matches 197; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

QY 82 CAAAGAGTCGGAGCTGGAGCTGGGCGTGTGTGGCGGCTGTGGCCAGCATGTTGGT 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172624 CAAAGTACCGGTACACTTCTTGGCGCGAGAGTGTCATTCTTCTTCATTGGCT 172683
QY 142 AAAGTAGTGTGAGTTCGAATGGCCATTGCTGTCTTGTGGGTGATTAAATGGT 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172684 CAAAGTAAAGACAACTTGTGCGAGTAGGTGAGGCGCATTTCTTGGGCGAGTCTTGGT 172743
QY 202 TCTAAATCGCTCAATCGATCGATCAGCAGGATTAATAAGCTAAACCGAGTTTGGAA 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172744 GGAACAATCGCGCAAGTAGATGAGCAGGATGAAGACTTCTAGAACCTAACATACAA 172803
QY 262 AAGTAAAGCAGGCGCAAGTGGCGTAAATCCAGATACAGCATTAAGTATAGT 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172804 AGAGCTTTAGAACTGCACTACGCGTA---GTACATTAATAATGGCCCAATCCAGATTAAC 172860
QY 322 GTTGAGCAGTGGCTACTTACACAGCGTTACAAATTAAGCAAGAGCGTCCGCAATATTTGT 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172861 GGCATATCATGTTAGCGTACACCTTAATAAACTTATAGAAACAGTGCAGGTCAATATTGC 172920
QY 382 CGAGATTTTACGAAAAAGCGATGATTCAGGCGCAGAGCAAGAGATTTCAGGCACATGCA 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172921 CGTGAATACACTCAACAGTTATTAATAGGCGGAAACACAAAAACATATGTGTAATGCA 172980
QY 442 TGCCGCGAACCGATGGTCTGTGGCAAGT 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172981 TGCCGTCAACCTGATGGCAATGGCAAGT 173009

```

```

LOCUS      AF031534      416 bp      DNA      linear      BCT 09-JAN-1998
DEFINITION Rickettsia cooley1 17 kDa antigen gene, partial cds.
ACCESSION  AF031534
VERSION    AF031534.1 GI:2654012
KEYWORDS   Rickettsia cooley1.
SOURCE     Rickettsia cooley1.
ORGANISM   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
            Rickettsiaceae; Rickettsiidae; Rickettsia.
REFERENCE  1 (bases 1 to 416)
AUTHORS   Billings,A.N., Tellow,G.J. and Walker,D.H.
TITLE     Molecular characterization of a novel spotted fever group
            rickettsial species from Ixodes scapularis in Texas
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 416)
AUTHORS   Billings,A.N., Tellow,G.J. and Walker,D.H.
TITLE     Submitted (28-OCT-1997) Pathology, UTHB at Galveston, 301
            University Blvd., Galveston, TX 77555-0609, USA
FEATURES   location/Qualifiers
            source
            1..416
            /organism="Rickettsia cooley1"
            /db_xref="taxon:69410"
            <1..415
            /codon_start=2
            /transl_table=11
            /product="17 kDa antigen"
            /protein_id="AB95267.1"
            /db_xref="GI:2654013"
            /translation="GSYIGPRGKNNKGGTLLGAGGALLGSQFGKGLVGVGVA
            LGAVLGGIAGMDQDRRLAETSORLEAPSGSSTEMRPDNGNYVPNNY
            RNSGYCREYQTVYIGKKQKAYGNACRQPEQ"
BASE COUNT 133 a 80 c 108 g 95 t
ORIGIN
Query Match      14.1%: Score 68.4; DB 1; Length 416;
Best Local Similarity 49.5%; Pred. No. 4.1e-08;
Matches 205; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 51 TTTAGTGGCTGTGCCGAGAACTTAACTCGTCAAGAAGTCGAGCTGCGAGCTGT 110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TTCAATATATCGGTCACGCGGTATGAATAAACAGCACAGCACTTCTTGGCGGTGC 63
QY 111 TGTGGCGGTGCTGCTGCCAGCTGTTGGTAAAGTAGTGTGATTCGAATGGCCAT 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 CGGCGGTGCATTTCTTGGCTCTCAATTCGTAAGGTAAGGACAGCTTCTCGAGTAGG 123
QY 171 TGGTGTGCTGTTTGGGTGATTAATGGTTCTAAATCGTCATCATGATGATCAGCA 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 TGTAGTGCATTAATCTTGACAGAGTCTTGGTGGCAAAATCGGTGCAGATGATGACGA 183
QY 231 GGATAAATTAAGCTAAACAGAGTTTGGAAAGTAAGGTAAGGCAAGTGCACGCTTG 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 GATAGAGAGCTTGCACAACTCACTCACAAGAGCTTTAGAG---CACCTCTACGCG 240
QY 291 GCGTAATCCAGATACAGAGCAATAGTTATAGTGTGAGCCAGTGGCTACTTAACAGCGTTA 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TAGTATGACAGAAATGCGGTATCTGTAAAGGCAATTCAGTTATATCAACCTTAATA 300
QY 351 CAATTAACAGAGCGTGGCCAGCAATATTTGCGAATTTTCAGAAAAAGCGATGATTCG 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 AACTTATAGAAATTAAGCACTGTCATATTTGCCGAGATATATCTCAACACAGTTGTAATAGC 360
QY 411 AGGCGAGACAGAGATTTACGCGACTGCATGCGCGGCAACCGGATGCTGTTG 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 CGGAAACCAACAAACAAACATACGCTAATGCTATGCTGCCCAACCTGACGAACAATG 414

```

RESULT 5  
AF031534

RESULT 6  
R1RANT17KB 535 bp DNA linear BCT 26-APR-1993  
LOCUS R1RANT17KB  
DEFINITION R.typhi 17k genus-common antigen gene, complete cds.  
ACCESSION M28481

VERSION M28481.1 GI:152459  
KEYWORDS antigen.  
SOURCE R.typhi DNA.  
ORGANISM Rickettsia typhi  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Rickettsiaceae; Rickettsia; Rickettsia: typhus group.  
REFERENCE 1 (bases 1 to 535)  
AUTHORS Anderson, B.E. and Tzianabos, T.  
TITLE Comparative sequence analysis of a genus-common rickettsial antigen  
JOURNAL J. Bacteriol. 171, 5199-5201 (1989)  
MEDLINE 89359171  
COMMENT Draft entry and printed copy of sequence for [1] kindly submitted  
FEATURES  
Source  
Location/Qualifiers  
1..535  
/organism="Rickettsia typhi"  
/db\_xref="taxon:785"  
-35\_signal  
1..6  
-10\_signal  
21..27  
mrna  
37..253  
/product="17 kd mRNA"  
56..535  
/note="17 kd antigen precursor"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA26377.1"  
/db\_xref="GI:152460"  
/translation="MKLLSKVMTIALATSMLOACNGPGMKNKGTGLLGAGGALLG  
RNPONGHGVYTPNKTYRNSTGOCREYTGTVIGKQOYTYNACRQPDGMQVYN"  
sig\_peptide  
56..115  
/note="17 kd antigen signal peptide"  
mat\_peptide  
116..532  
/product="17 kd antigen"  
BASE COUNT 182 a 93 c 122 g 138 t  
ORIGIN  
Query Match  
Best Local Similarity 13.5%; Score 65.8; DB 1; Length 535;  
Matches 194; Conservative 0; Mismatches 192; Indels 3; Gaps 1;  
OY 82 CAAGAAGTCGGAGCTGCGACCTGGGCGTGTGTCGGCGGTGCTGCGCCAGCTGTTGGT 141  
DB 140 CAAGGACTGGAACACTTCTTGGGCGCGCGGCGGCACTTCTGCTTCAATTCGCT 199  
OY 142 AAGGTAAGTGTGCGAGTTCGCAATGCGCATGCGTGTGCTGTTGGGCGATTCGCT 201  
DB 200 CACGCTAAGGACAACTTGTGCGAGTAGTGTAGCGGCACTTCTGAGGCACTTCTTGT 259  
OY 202 TCTAATAATCGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261  
DB 260 GGAACAATCGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 319  
OY 262 AAGGTAAGTGTGCGAGTTCGCAATGCGCATGCGTGTGCTGTTGGGCGATTCGCT 321  
DB 320 AGAGCTTTAGAAC---GCTCTAGCGGAGTAGTAATGAAATGCGCAATTCAGATAT 376  
OY 322 GTTGAGCCAGTGCCTACTTACCACTTACCAATGAAAGAGCGTCCGACCATATTTGT 381  
DB 377 GGCATTCATGCTTACCTTACCACTTACCAATGAAAGAGCGTCCGACCATATTTGT 436  
OY 382 CGAATAATTCAGCAAAAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 441  
DB 437 CGTGAATACACTCAAAAGAGTGTATAGGGGGAACAAACAAACAAACATATGGAATGCA 496  
OY 442 TGCCGCAACCGGATGCTGTTGCAAGT 470  
DB 497 TGCCGCAACCGGATGCTGTTGCAAGT 525

LOCUS R17K6CA 537 bp DNA linear BCT 04-FEB-1999  
DEFINITION Rickettsia japonica 17K genus-common antigen gene, complete cds.  
ACCESSION D16515  
VERSION 17K  
KEYWORDS 17K genus-common antigen; 17K dalton protein.  
SOURCE Rickettsia japonica (sub-species: JH) DNA.  
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Rickettsiaceae; Rickettsia; Rickettsia: spotted fever group.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Yoshida, Y.  
TITLE Direct Submission  
JOURNAL Submitted (05-JUN-1993) Yoshida, Y., Kanagawa Prefectural  
Public Health Laboratory, 52-2 Nakao-cho, Asahi-ku, Yokohama,  
Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)  
Furuya, Y., Katayama, T., Yoshida, Y. and Kaiho, I.  
Specific amplification of Rickettsia japonica DNA from clinical  
specimens by PCR  
J Clin. Microbiol. 33 (2), 487-489 (1995)  
FEATURES  
Source  
Location/Qualifiers  
1..537  
/organism="Rickettsia japonica"  
/sub\_species="JH"  
/db\_xref="taxon:35790"  
/note="propagated in Vero cells"  
58..537  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAA03965.1"  
/db\_xref="GI:1435056"  
/translation="MKLLSKVMTIALATSMLOACNGPGMKNKGTGLLGAGGALLG  
RNPONGHGVYTPNKTYRNSTGOCREYTGTVIGKQOYTYNACRQPDGMQVYN"  
BASE COUNT 176 a 92 c 129 g 140 t  
ORIGIN  
Query Match  
Best Local Similarity 13.5%; Score 65.8; DB 1; Length 537;  
Matches 194; Conservative 0; Mismatches 192; Indels 3; Gaps 1;  
OY 82 CAAGAAGTCGGAGCTGCGACCTGGGCGTGTGTCGGCGGTGCTGCGCCAGCTGTTGGT 141  
DB 142 CAAGTACGACAACTTCTTGGGCGCGCGGCGGCACTTCTGCTTCAATTCGCT 199  
OY 142 AAGGTAAGTGTGCGAGTTCGCAATGCGCATGCGTGTGCTGTTGGGCGATTCGCT 201  
DB 200 CACGCTAAGGACAACTTGTGCGAGTAGTGTAGCGGCACTTCTGAGGCACTTCTTGT 259  
OY 202 TCTAATAATCGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261  
DB 260 GGAACAATCGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 319  
OY 262 AAGGTAAGTGTGCGAGTTCGCAATGCGCATGCGTGTGCTGTTGGGCGATTCGCT 321  
DB 320 AGAGCTTTAGAAC---GCTCTAGCGGAGTAGTAATGAAATGCGCAATTCAGATAT 376  
OY 322 GTTGAGCCAGTGCCTACTTACCACTTACCAATGAAAGAGCGTCCGACCATATTTGT 381  
DB 377 GGCATTCATGCTTACCTTACCACTTACCAATGAAAGAGCGTCCGACCATATTTGT 436  
OY 382 CGAATAATTCAGCAAAAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 441  
DB 437 CGTGAATACACTCAAAAGAGTGTATAGGGGGAACAAACAAACAAACATATGGAATGCA 496  
OY 442 TGCCGCAACCGGATGCTGTTGCAAGT 470  
DB 499 TGCCGCAACCGGATGCTGTTGCAAGT 527

LOCUS	R1RAN1717KA	539 bp	DNA	linear	BCT 26-APR-1993
DEFINITION	R. conorii 17K genus-common antigen gene, complete cds.				
ACCESSION	M28480				
VERSION	M28480.1	GI:152457			
KEYWORDS	antigen.				
SOURCE	R. conorii DNA.				
ORGANISM	Rickettsia conorii				
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiaseae; Rickettsia; spotted fever group. 1 (bases 1 to 539)				
AUTHORS	Anderson, B.E. and Tzianabos, T.				
TITLE	Comparative sequence analysis of a genus-common rickettsial antigen gene				
JOURNAL	J. Bacteriol. 171, 5199-5201 (1989)				
MEDLINE	89359171				
COMMENT	Draft entry and printed copy of sequence for [1] kindly submitted by B. Anderson, 27-OCT-1989.				
FEATURES	Location/Qualifiers				
Source	1..539				
-35_signal	/organism="Rickettsia conorii"				
-10_signal	/db_xref="taxon:781"				
mrna	1..6				
cds	21..27				
	38..>539				
	/product="17 kd mrna"				
	60..539				
	/note="17 kd antigen precursor"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="AAA26376.1"				
	/db_xref="GI:152458"				
	/translation="MKDLISKIMIALATSMLOACNPGPNKKGSTGTLILGGAGGALLG				
	SOFKSGGLGVGVGALLGVALVGGAGDEDDRLAELTSGRALETPSGSNVEM				
	RNPNGNYGVYTPPKTKRNSPGCYCRVYTGTVVIGRQKRAYGNACHPGQWQVNA"				
	60..119				
sig_peptide	/note="17 kd antigen signal peptide"				
mat_peptide	120..536				
	/product="17 kd antigen"				
BASE COUNT	179 a 125 g 142 t				
ORIGIN	93 c				
Query Match	13.3%: Score 64.6; DB 1; Length 539;				
Best Local Similarity	49.4%: Pred. No. 4.7e-07;				
Matches 197; Conservative	0; Mismatches 199; Indels 3; Gaps 1;				
0Y	72	CTTTGTCGCTGAAGAGTCGAGCGACGTCGACGCGGGCTGTTGGCGGCTGTCGGCA	131		
Db	134	CATGATATAACCAAGGTACAGGAACCTTCTTGCGCGTCTGGCGGCATTACTGCTTC	193		
0Y	132	GCTGTGCTGAAGGTAGTGTGTCGATTCGATTCGCAATGGCCATTGGTGTGCTGTTTGGTGG	191		
Db	194	TCAATTGCGTAAAGGCCAAGGACACTTGTGGATGATGAGTGCACTTACTTGGAGC	253		
0Y	192	ATTAATTGGTTCCTAAATCGGTCATTCGATGCATGCAGCAGCATTAATPAAACTTAACCA	251		
Db	254	AGTCTTGTGCGTACAAATCGGTGCGAGTATGATGAACAGGATTAAGAACTTGCGAGACT	313		
0Y	252	GAGTTTGGAAAAAGGTAAGGAGGCGGCAAGTACAGCTGGGTATCCAGATACAGGCA	311		
Db	314	TACCTCACAGAGAGCTTTAGAAAC--AGCTCCCTTAGTGTAGTAAACGTAAGTAGGCGTAA	370		
0Y	312	TAGTATATAGTGTGAGCCAGTCGCTACTTTACACAGCGTTACATATAAGCAAGAGCTCGCA	371		
Db	371	TCCGGATTAAGCGCACTTACGCTTACGTAAACACTATATAAACTTATAGAAATAGCACATCG	430		
0Y	372	GCAATATGTCGAGAAATTCAGACAAAGGCGATGATTCAGGCGCAGAAAGCAAGAGATTTA	431		
Db	431	TCAATATTGCGGTGAGTACACTCAACACAGTTGTAATAGGCGGAAAAACAACAAAAGACATA	490		
0Y	432	CGGCACTGATCCCGGCAACCGGATGCTGTTGGCACT	470		
Db	491	CGGTAAATGATCCGCGCAACTGAGCGGCAATAGGCAACT	529		

LOCUS	AE008675	10127 bp	DNA	linear	BCT 14-SEP-2001
DEFINITION	Rickettsia conorii Malish 7, section 107 of 114 of the complete genome.				
ACCESSION	AE008675	AE006914			
VERSION	AE008675.1	GI:15620413			
KEYWORDS					
SOURCE	Rickettsia conorii.				
ORGANISM	Rickettsia conorii				
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.				
AUTHORS	Ogata, H., Audic, S., Barbe, V., Ariguenave, F., Fournier, P.E., Raoult, D. and Claverie, J.M.				
TITLE	Selfish DNA in protein-coding genes of Rickettsia				
JOURNAL	Science. 290 (3490), 347-350 (2000)				
MEDLINE	20485652				
PUBMED	11030655				
REFERENCE	2 (sites)				
AUTHORS	Ogata, H., Audic, S. and Claverie, J.-M.				
TITLE	Selfish DNA and the origin of genes				
JOURNAL	Science 291 (5502), 252-253 (2001)				
REFERENCE	3 (bases 1 to 10127)				
AUTHORS	Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.				
TITLE	Mechanisms of Evolution in Rickettsia conorii and R. prowazekii				
JOURNAL	Science. 293 (5537), 2093-2098 (2001)				
PUBMED	11537893				
REFERENCE	4 (bases 1 to 10127)				
AUTHORS	Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 1385, France				
COMMENT	A public version of R. conorii genome database is accessible at <a href="http://igs-server.cirs-mrs.fr/">http://igs-server.cirs-mrs.fr/</a> . The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.				
FEATURES	Location/Qualifiers				
source	1..10127				
gene	/organism="Rickettsia conorii"				
	/strain="Malish 7"				
	/db_xref="taxon:781"				
	111..890				
	/gene="map"				
	/note="RC1276"				
	111..890				
	/gene="map"				
	/codon_start=1				
	/transl_table=1				
	/product="methionine aminopeptidase [EC:3.4.11.18]"				
	/protein_id="AAL03814.1"				
	/db_xref="gi:15620414"				
	/translation="MTIKIRHGKQFIRKMRAGKLAETLFTIDVHKNVNTYNSLNLCHNFTISYVAVPAPLNKGGPKSICSTINHYCHGIPNDKRLKGDIVNIDVYIILGGYVILGGSITMYVGVDAIKPKRIQVTPAMMKGIYVVRPAKLGIDIGHAITSYAEKHNYSVAVDYTHGIGRVEHDEPSILNCRSGTGLTEEGHFFPYEPIINAGNTDTILSKLDGWITVTRDKSLSAQFEHTIGVTKGFEFTLSPKKLDPPY"				
	complement(977..1011)				
	/note="REP03, repeated element"				
	complement(1064..1083)				
	/note="REP02, RS3-like repeat"				
	1141..1452				
	/gene="RC1277"				
	1141..1452				
	/gene="RC1277"				

gene  
1457. 1825  
/gene="RC1278"  
1457. 1825  
/gene="RC1278"  
/codon\_start=1  
/transl\_table=1  
/product="similarity to DNA repair protein (Radc)"  
/protein\_id="AAL03815.1"  
/db\_xref="GI:15620415"  
/translation="MLENNLEFSENFDEIDTPYIGHRRKRYERYVAAGAEYSDEE  
LEVMFSAIPRDKIFLAKKLDFDITDILNDKRELISLIGTBNENYINFALELL  
N"  
1457. 1825  
/gene="RC1278"  
1457. 1825  
/gene="RC1278"  
/codon\_start=1  
/transl\_table=1  
/product="similarity to DNA repair protein (Radc)"  
/protein\_id="AAL03816.1"  
/db\_xref="GI:15620416"  
/translation="MKQKITKNINASSWSDIDYTKVMGNMRLEQFRIELANKNL  
IADEYLSGCTIDQAAVPREIIRALFNEASSLIVHHPSGSEPPSSAHIMHNKIV  
EKQCTINIVNLCSYEK"  
1812. 2795  
/gene="mray2"  
/note="RC1279"  
1812. 2795  
/gene="mray2"  
/codon\_start=1  
/transl\_table=1  
/product="undecaprenyl-phosphate  
alpha-N-acetylglucosaminyltransferase [EC:2.4.1.-]"  
/protein\_id="AAL03817.1"  
/db\_xref="GI:15620417"  
/translation="MNINSYLFTEFSPTSAITTYILTYLEPTGLVDIPSNRSHD  
KIRPGGGAIVPVYVIALISGPEYITTSNENYSKILPLILAIASISFLDDLKAVIL  
RLIVHLIOAFAPIELFQSLNSVHLIYLSIALVSGFINIVNEMGIDIGMSVSIH  
LSTSMITFLQPTIINNPTFTTSVNVILLEGSGFLIFMHPAKIFLDGVSISGF  
LEPCLILLALISLTFACAIASLDYITTDVAVITLIRLNRKRIWQPLKHPFKAV  
QAKSKQVQVSIIVACNIFLIMISVTSLHFEVVISIMLAQVITLVQSILK"  
2792. 3115  
/gene="RC1280"  
2792. 3115  
/gene="RC1280"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAL03818.1"  
/db\_xref="GI:15620418"  
/translation="MKLIKEDIYVHIFPAKHPLPEPIMINMKIYGFNEFKALP  
LKITYTYKKKIKINTLFLQADENATAELPYODITLERIKIYILGFAIHOMNTEFK  
GKRSI"  
3113. 3353  
/note="REP01, repeated element"  
3113. 3353  
/note="REP02, RS3-like repeat"  
3113. 3353  
/note="REP03, repeated element"  
3113. 3353  
/note="REP04, repeated element"  
3113. 3353  
/note="REP05, repeated element"  
3113. 3353  
/note="REP06, repeated element"  
3113. 3353  
/note="REP07, repeated element"  
3113. 3353  
/note="REP08, repeated element"  
3113. 3353  
/note="REP09, repeated element"  
3113. 3353  
/note="REP10, repeated element"  
3113. 3353  
/note="REP11, repeated element"  
3113. 3353  
/note="REP12, repeated element"  
3113. 3353  
/note="REP13, repeated element"  
3113. 3353  
/note="REP14, repeated element"  
3113. 3353  
/note="REP15, repeated element"  
3113. 3353  
/note="REP16, repeated element"  
3113. 3353  
/note="REP17, repeated element"  
3113. 3353  
/note="REP18, repeated element"  
3113. 3353  
/note="REP19, repeated element"  
3113. 3353  
/note="REP20, repeated element"  
3113. 3353  
/note="REP21, repeated element"  
3113. 3353  
/note="REP22, repeated element"  
3113. 3353  
/note="REP23, repeated element"  
3113. 3353  
/note="REP24, repeated element"  
3113. 3353  
/note="REP25, repeated element"  
3113. 3353  
/note="REP26, repeated element"  
3113. 3353  
/note="REP27, repeated element"  
3113. 3353  
/note="REP28, repeated element"  
3113. 3353  
/note="REP29, repeated element"  
3113. 3353  
/note="REP30, repeated element"  
3113. 3353  
/note="REP31, repeated element"  
3113. 3353  
/note="REP32, repeated element"  
3113. 3353  
/note="REP33, repeated element"  
3113. 3353  
/note="REP34, repeated element"  
3113. 3353  
/note="REP35, repeated element"  
3113. 3353  
/note="REP36, repeated element"  
3113. 3353  
/note="REP37, repeated element"  
3113. 3353  
/note="REP38, repeated element"  
3113. 3353  
/note="REP39, repeated element"  
3113. 3353  
/note="REP40, repeated element"  
3113. 3353  
/note="REP41, repeated element"  
3113. 3353  
/note="REP42, repeated element"  
3113. 3353  
/note="REP43, repeated element"  
3113. 3353  
/note="REP44, repeated element"  
3113. 3353  
/note="REP45, repeated element"  
3113. 3353  
/note="REP46, repeated element"  
3113. 3353  
/note="REP47, repeated element"  
3113. 3353  
/note="REP48, repeated element"  
3113. 3353  
/note="REP49, repeated element"  
3113. 3353  
/note="REP50, repeated element"  
3113. 3353  
/note="REP51, repeated element"  
3113. 3353  
/note="REP52, repeated element"  
3113. 3353  
/note="REP53, repeated element"  
3113. 3353  
/note="REP54, repeated element"  
3113. 3353  
/note="REP55, repeated element"  
3113. 3353  
/note="REP56, repeated element"  
3113. 3353  
/note="REP57, repeated element"  
3113. 3353  
/note="REP58, repeated element"  
3113. 3353  
/note="REP59, repeated element"  
3113. 3353  
/note="REP60, repeated element"  
3113. 3353  
/note="REP61, repeated element"  
3113. 3353  
/note="REP62, repeated element"  
3113. 3353  
/note="REP63, repeated element"  
3113. 3353  
/note="REP64, repeated element"  
3113. 3353  
/note="REP65, repeated element"  
3113. 3353  
/note="REP66, repeated element"  
3113. 3353  
/note="REP67, repeated element"  
3113. 3353  
/note="REP68, repeated element"  
3113. 3353  
/note="REP69, repeated element"  
3113. 3353  
/note="REP70, repeated element"  
3113. 3353  
/note="REP71, repeated element"  
3113. 3353  
/note="REP72, repeated element"  
3113. 3353  
/note="REP73, repeated element"  
3113. 3353  
/note="REP74, repeated element"  
3113. 3353  
/note="REP75, repeated element"  
3113. 3353  
/note="REP76, repeated element"  
3113. 3353  
/note="REP77, repeated element"  
3113. 3353  
/note="REP78, repeated element"  
3113. 3353  
/note="REP79, repeated element"  
3113. 3353  
/note="REP80, repeated element"  
3113. 3353  
/note="REP81, repeated element"  
3113. 3353  
/note="REP82, repeated element"  
3113. 3353  
/note="REP83, repeated element"  
3113. 3353  
/note="REP84, repeated element"  
3113. 3353  
/note="REP85, repeated element"  
3113. 3353  
/note="REP86, repeated element"  
3113. 3353  
/note="REP87, repeated element"  
3113. 3353  
/note="REP88, repeated element"  
3113. 3353  
/note="REP89, repeated element"  
3113. 3353  
/note="REP90, repeated element"  
3113. 3353  
/note="REP91, repeated element"  
3113. 3353  
/note="REP92, repeated element"  
3113. 3353  
/note="REP93, repeated element"  
3113. 3353  
/note="REP94, repeated element"  
3113. 3353  
/note="REP95, repeated element"  
3113. 3353  
/note="REP96, repeated element"  
3113. 3353  
/note="REP97, repeated element"  
3113. 3353  
/note="REP98, repeated element"  
3113. 3353  
/note="REP99, repeated element"  
3113. 3353  
/note="REP100, repeated element"  
3113. 3353  
/note="REP101, repeated element"  
3113. 3353  
/note="REP102, repeated element"  
3113. 3353  
/note="REP103, repeated element"  
3113. 3353  
/note="REP104, repeated element"  
3113. 3353  
/note="REP105, repeated element"  
3113. 3353  
/note="REP106, repeated element"  
3113. 3353  
/note="REP107, repeated element"  
3113. 3353  
/note="REP108, repeated element"  
3113. 3353  
/note="REP109, repeated element"  
3113. 3353  
/note="REP110, repeated element"  
3113. 3353  
/note="REP111, repeated element"  
3113. 3353  
/note="REP112, repeated element"  
3113. 3353  
/note="REP113, repeated element"  
3113. 3353  
/note="REP114, repeated element"  
3113. 3353  
/note="REP115, repeated element"  
3113. 3353  
/note="REP116, repeated element"  
3113. 3353  
/note="REP117, repeated element"  
3113. 3353  
/note="REP118, repeated element"  
3113. 3353  
/note="REP119, repeated element"  
3113. 3353  
/note="REP120, repeated element"  
3113. 3353  
/note="REP121, repeated element"  
3113. 3353  
/note="REP122, repeated element"  
3113. 3353  
/note="REP123, repeated element"  
3113. 3353  
/note="REP124, repeated element"  
3113. 3353  
/note="REP125, repeated element"  
3113. 3353  
/note="REP126, repeated element"  
3113. 3353  
/note="REP127, repeated element"  
3113. 3353  
/note="REP128, repeated element"  
3113. 3353  
/note="REP129, repeated element"  
3113. 3353  
/note="REP130, repeated element"  
3113. 3353  
/note="REP131, repeated element"  
3113. 3353  
/note="REP132, repeated element"  
3113. 3353  
/note="REP133, repeated element"  
3113. 3353

Query Match	Best Local Similarity	Score	DB 1:	Length
Matches 197; Conservative	0;	Mismatches 199;	Indels	3; Gaps 1;
192	ATTAAATGGTCTAAATATCGGTCATATCGATGATCAGCAGATAAATAAAGCTAAACCA	251		
Db	8208	CATGAATTAACAAAGAGTACAGGAACACTTCTTGGCGGTGCGGGCGGCATTACTTGTTGC	8267	
132	GCTGTTTGGTAAAGGTAGTGTGAGTGTGCAATAGCCATTTGGTGGTGGCTGTTTGGGTGG	191		
Db	8268	TCAATTCGTAAGGCGAAAGACAGCTTTGTGAGTAGGTGATGATTAACCTTGAGAC	8327	
192	ATTAAATGGTCTAAATATCGGTCATATCGATGATCAGCAGATAAATAAAGCTAAACCA	251		

QY	202	TCCTAAATTCGCTCATTCATGATGCATCAGCATATAAAATGAAGCTAAACCAAGATTGGCA	261
Db	264	GGACAAATCGGTGCAGAGTATGATGAACAGATAGAAAGACTTGGCAGAGCTTACTCTCACAG	323
QY	262	AAGGTAAGAAAGCAGAGGCAAGTGCACAGCGTTGGGGGTATATCCAGATACAGGCATATAGTTATAGT	321
Db	324	AGAGCTTTAGAAAC---AGCTCTTAGTGTGTAGTACGTAAGATGGCGTTAAATCCGGATTAAC	380
QY	322	GTTGAGCCAGTGCCTACTTACTTACCAGCGTTTACCAATTAAGCAAGAGCGCTGGCAGCAATATTGT	381
Db	381	GGCAATTACGGTTTACGTATACACCTAATTAATTAACCTTATATACACATCGTCAATATTGCG	440
QY	382	CGAGATTTCAGCAAAAGCGCATGATTGCAGGCGCAAGACAAAGATTACGGCATGCA	441
Db	441	CGTGAGTCACTCAAAACAGTGTATATAGCGGAAACACAAAGCATACGTAATATGCA	500
QY	442	TGCGGCAACCGAGTGCCTGGCAAGT	470
Db	501	TGCGGCAACCTGAAGCAATGGCAAGT	529
RESULT	11		
	R1RANTR		
LOCUS	R1RANTR	620 bp	DNA
DEFINITION	R.rickettsii 17k antigen gene, complete cds.	linear	BCT 26-Apr-1993
ACCESSION	M16486		
VERSION	M16486.1	GI:152467	
KEYWORDS	antigen.		
SOURCE	R.rickettsii (strain Sheila Smith) DNA, clone pSC2.		
ORGANISM	Rickettsia rickettsii		
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group. 1 (bases 1 to 620)		
AUTHORS	Anderson, B.E., Regnery, R.L., Carlone, G.M., Tzianabos, T., McDade, J.E., Fu, Z.Y. and Bellini, W.J.		
TITLE	Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia rickettsii		
JOURNAL	J. Bacteriol. 169, 2385-2390 (1987)		
MEDLINE	87222152		
COMMENT	Draft entry and computer-readable sequence for [1] kindly provided by B.E.Anderson, 23-Oct-1987.		
FEATURES	-35 and -10 regions are located at positions 12-17 and 33-38.		
	Location/Qualifiers		
	1..620		
	/organism="Rickettsia rickettsii"		
	/db_xref="taxon:783"		
	70..126		
	/note="17kd antigen signal peptide"		
	70..549		
	/note="17kd antigen precursor"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA26381.1"		
	/db_xref="GI:152468"		
	/translation="MKLLSKIMTIALTSMLOACNGPGNNKOGTGLLGAGAGALLG		
	SOPEKSGQGLVGVGALGAVLGGQICGAMDEDDRLAELTSGRALETPSGSVEM		
	RNPNGNYGVYTPPKTKRYRNSYGGYCREYTGTVVIGKQKAYGADGCRPDQMOMVN"		
	127..546		
	/product="17kd antigen"		
	206 a	105 c	142 g
	167 t		
	BASE COUNT		
	Origin	Unreported.	
	Query Match	13.2%:	Score 64.2; DB 1; Length 620;
	Best Local Similarity	49.6%:	Pred. No. 6.2e-07;
	Matches 193; Conservative	0; Mismatches 193;	Indels 3; Gaps 1;
QY	82	CAAGCAAGTCGACGCTGCACACTGGGGCTGTGTGGCGGTGTGGCGACACTGTTGGT	141
Db	154	CAAGTACAGCAACACTCTTGGCGGTCTGGCGGCGCATTAACCTTGCTTCATATGCGT	213
QY	142	AAAGCTAGTGGTCGACGTTGCACATGGCCATTGGTGCTGTTTGGGTGCAATTAATGGT	201

Query Match 12.88; Score 62.2; DB 1; Length 434

```

FEATURES
    source
        location/Qualifiers
            1..434
                /organism="male-killing Rickettsia from Adalia bipunctata"
                /specific_host="Adalia bipunctata"
                /db_xref="taxon:38028"
                /country="United Kingdom: Cambridge"
                /note="Isolated from 2spot ladybird"
            1..20
                /note="PCR primer R1"
            <1..>434
                /codon_start=1
                /transl_table=11
                /product="17 kDa antigen"
                /protein_id="CAB96382.1"
                /db_xref="GI:8920295"
                /db_xref="SPTREMBL:Q9K2N6"
    CDS

```



```

primer_bind
/note="PCR primer R2"
415..434
BASE COUNT      132 a      88 c      114 g      100 t
ORIGIN
```

Query Match	12.8%	Score 62.2	DB 1	Length 434
Best Local Similarity	50.1%	Pred. No. 2.1e+06		
Matches 182, Conservative	0	Mismatches 178	Indels 3	Gaps 1

OY	102	TGGGGCTTTGTGGCGGTGTGTGCTGGCCAGCTGTTTGTAAAGGTAGTGGTCAGTTGC	161
Db	75	TGGCGGTCCCGCGCGGTGCATTACTTGGTTCTCAATTCGGCAAAAGTTAAAGGCCAATTGT	134
OY	162	AATGGCCATTGGTGTGCTGTGTTTGGGTGGATTTAATTGTTCTAAATCCGTCATTCGAT	221
Db	135	CGGAGTAGGTAGTGGTCATTACTTGGAGACAGTTCTTGTTGGACAAATCCGTCGAGGTAT	194
OY	222	GGATCAGCAGGATTTAAATTAAGCTTAACCCAGAGTTTGGAAAAGTTAAAGCAGGGCAAGT	281
Db	195	GGATGACGAGGATAGAAACATTGCTGAACTCACCTCTCAAAAGAGCTTTAAGAAGC---AGC	251
OY	282	GACACGTTGGGGTATTCAGATACAGGCAATATGTTATAGTGTGTAGCCAGTCCGACTTA	341
Db	252	TTCCTAGTGTGTAGTAACGTAGAATGGCGGTAAATCCGATAAACGGCAATCATGGCTACGTAAAC	311
OY	342	CCAGCGTAACTAATAGCAAGAAGCGTCCGCAGCAATATTGTGAGAAATTTACGCAAAAAGCC	401
Db	312	ACCTAATATAAACTTATAGAAATAGCACAGGCGCAATATTGTCCTGTGAGTACACTCAAAACAGT	371
OY	402	GATGATTCGAGGGCGAAGAACAGATTTTACGGCACTGATGCCGCAACCGGATGCTCG	461
Db	372	TGTAATAGCGGAAAAACAACAAAAAGCATACGGTAACTCATGATCCGCCCAACCTGACGAGACA	431
OY	462	TTG	464
Db	432	ATG	434

RESULT	14
Locus	MRI269518
Definition	MRI269518      434 bp    DNA           linear   BCF_27-JAN-2001 Male-killing Rickettsia from Adalia bipunctata partial 17kDa antigen gene (Denmark: Ribe).
Accession	AJ269518
Version	AJ269518.1 GI:8920296
Keywords	17 kDa antigen. male-killing Rickettsia from Adalia bipunctata.
Source	Bacteria: Proteobacteria; alpha subdivision: Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia.
Organism	1 (bases 1 to 434) von Der Schulenburg,J.H., Habig,M., Sloggett,J.J., Webberley,K.M., Bertland,D., Hurst,G.D. and Majerus,M.E. Incidence of male-killing rickettsia spp. (alpha-Proteobacteria) in the ten-spot ladybird beetle adalia decemmaculata L Appl. Environ. Microbiol. 67 (1), 270-277 (2001)
Reference	2 (bases 1 to 434) Schulenburg,H. Direct Submission Submitted (03-APR-2000) Schulenburg H., Department of Evolutionary Biology, Institut fuer Spezielle Zoologie, Huefferstr. 1, 48149 Munster, GERMANY
Title	Location/Qualifiers
Journal	1..434
Medline	/organism="male-killing Rickettsia from Adalia bipunctata"
Authors	/specific_host="Adalia bipunctata"
Reference	/db_xref="taxon:38028"
Journal	/country="Denmark: Ribe"
Features	/note="Isolated from 2spot ladybird"
source	

```

primer_bind      1. 20
                  /note="PCR primer R1"
CDS              <1..>434
                  /codon_start=1
                  /transl_table=1
                  /product="17kDa antigen"
                  /protein_id="CAB96383.1"
                  /db_xref="GI:8920297"
                  /db_xref="SPTREMBL:O9K2N6"
                  /translation="ALATSMQLQACTGPGGNKKQGTITLIGAGGALLGSGQFGRKGGQ
VGVGVGALLGVALIGGQIGAGMDEQDRRLAELTSDALEAAPSQSVNEMRNPDNGNHG
VTPPKTYRNTSGQYCRXYTQFVIVIGGKQAKAYGNACRQPEDQ"
primer_bind      415..434
                  /note="PCR primer R2"
BASE COUNT      133 a      87 c      114 g      100 t
ORIGIN

```

Query Match	12.8%	Score 62.2	DB 1	Length 434
Best Local Similarity	50.1%	Pred. No. 2.1e-06		
Matches 182; Conservative	0	Mismatches 178;	Indels 3	Gaps 1

OY	102	TGGGGCTGTTGGCGGCTGTGCTGGCCAGCTGTTTGGTAAAGGATAGTCGAGTTGC	161
Db	75	TGGCGGTGCCGGCGGTGCATTACTTGTTTCTCAATTCGCAAAAGGTAAGGGCAACTTGT	134
OY	162	AATGGCCATTTGGTGGTGTGTTGGGTGGATTTAAATTTGGTTCTAAAAATCGTCAATTCGAT	221
Db	135	CGGATTAGTGTAAGTGGCTTACTTGGAGCACTTCTTGTTGGCAAAATCGGTGCGAGTAT	194
OY	222	GGATCAGCAGATTAATAATTAAGCTAAACCAGATTGGAAAGGTAAAGCAGGGCCAGT	281
Db	195	GGATGAGCAGGATAAAGAACTGTGTAATCTCACTCTCAAAAGAGCTTTGAAGAC--AGC	251
OY	282	GACAGTGGGGCTAATCCAGATACAGGCATAGTTATATAGTTGAGCCGTCGTACTTA	341
Db	252	TCCTAGTGGTACTACGTAGTAATGGCGTATCCGATTAACGGCAATCATGCTACGTAAC	311
OY	342	CCAGCGTTACATAATAGCAAGAGCGTCGCCAGCAATATGTCCAGAAATTTACAGCAAAAGGC	401
Db	312	ACCTAATATAAACTTTATAGAAATAGCAACAGGGCAATATTTTCCGTAGATACACTCAAAACAT	371
OY	402	GATGATTTGCAGGGCAGAAAGCAAGAGATTTTAGCGCACTTCGATCCCGCAACCGGATGCTCG	461
Db	372	TGTATATAGCGGAAACAACAACAAAGACATACGTTATGTCATCCGCCCAACCTCGAGAACCA	431
OY	462	TTG 464	
Db	432	ATG 434	

RESULT	15
LOCUS	AF195118
DEFINITION	Rickettsia felis 17 kDa genus-common antigen gene, complete cds.
ACCESSION	AF195118
VERSION	AF195118.1
KEYWORDS	GI:11066087
SOURCE	Rickettsia felis. Rickettsia felis Bacteria; Proteobacteria; alpha subdivision; Rickettiales; Rickettsiaceae, Rickettsiae; Rickettsia: spotted fever group. 1 (bases 1 to 346) Boyer/D.H., Stenos,J., Crocquet-Valdes,P., Moron,C.G., Popov,V.L., Vazala-Velazquez,J.E., Foll,L.D., Stothard,D.R., Azad,A.F. and Walker,D.H. Rickettsia felis: molecular characterization of a new member of the spotted fever group Int. J. Syst. Evol. Microbiol. 51 (Pt 2), 339-347 (2001)
JOURNAL	MEDLINE
PUBMED	21217364
REFERENCE	11321078
AUTHORS	2 (bases 1 to 346) Boyer/D.H., Stenos,J., Crocquet-Valdes,P.A., Foll,L.D. and Walker,D.H.

us-09-677-374-1.rge

Direct Submission  
Submitted 6/14/2007

Branch at Galveston, 301 University Blvd., Galveston, TX 77551-0609, USA

1999-2000, USA	Location/Qualifiers
1. .546	

```

/db_xref="taxon:42862"
65      54A
CDS

```

```

/coo001_start=1
/transl_table=11
/prod=1

```

```

/protein_id="AAG28452.1"
/db_xref="GI:11066088"

```

SQFGKKGQLVGVGVGALLGAV  
 RNP DNGNHGYVTPNKTVDNCS

```

BASE COUNT      183 a      96 c      126 g      141 t
ORIGIN
      240505060611GAGTGGGAGCAGMDEDDRRLAEITLTSOALFAPSGTSEW"
      RNPDGNHGIVTPNKRTYRNSTGOYCRETTOTVIVIGGKQKAYGNACRQPDGLMOVN"

```

Query Match  
Post r

Very match	12.28;	Score 59.4;	DB 1;	Length 546;
Best Local Similarity	48.88;	Pred. No. 1.3e-05;		
Matches 190; Conservative				

	82	CAGAGAATCGGACGTCCGACTGGGGCTCCTTTCCTTTTGCTTTT	Indels	3;	Gaps	1;
--	----	--	--------	----	------	----

D6  
149 CAAGGTACGGGAACATCTTTGCGCGCCGCGTTCCTGCCACAGCTGTTGGT 14

QY 142 AAAGTACTGCTGCAGTTGCAATTCGCGC 208

D6  
209 AAGGCAAGACACCTTCCCGCCTTTGGCTGATTAATTGGT 201

202 TCTAATAATCGGTCAATCGATTTCTTGGT 268

Ddb  
269 GGACAAATAGCCGCGTGGCTTTGCA 261

263 AAGCTAATGCGTGGTGAACCTCCTTCACA 328

320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

225 AGAGCTTTAGAGCAACACCCTA---GCGCACTAGCGTGAATGGCGTAATCCGCAATG 321

322 GTTGAGCCAGTGGCTACTTACCAGCGTTACATATAGCAAGAGCGTGGCCAGCAATGTTTATC 385

386 GGTAATCATGCTTACGTAAACACCTAATAAAACTTTATGAAGATAGCAGCCGGCGCAATATCT 381

382 CGAGATTTCAGCAAAAGCGATGATTCAGGGCAGAGCAATTC 445

[illegible]

442 TGCCGGAACCGATGGTCGTGGCAAGT 470  
|||||  
-----AAGCATTACGGTAATGCA 505

506 TGGCGCCAACTGACGACTATGGCACT 534

CC FORM -

Arch Ser-1000

Search completed: October 27, 2002, 17:39:09  
Job time : 1078.22 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:15:49 : Search time 973.679 seconds  
(without alignments)  
6736.842 Million cell updates/sec

Title: us-09-677-374-1

Perfect score: 486  
Sequence: 1 atgaacagaggtgttgc.....aagtcattcaacagaaaa 486

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_iny:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vtl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	4.3	661	9	BB517742
2	20	4.1	395	10	BE428593
3	20	4.1	673	10	BF643947
4	20	4.1	689	12	AZ731780
5	20	4.1	988	12	CNS06GLH
6	19	3.9	376	9	AJ003546
7	19	3.9	380	10	T59966
8	19	3.9	436	9	AM151916
9	19	3.9	461	12	AZ952777
10	19	3.9	545	9	A1965711
11	19	3.9	550	12	AZ887587
12	19	3.9	561	12	BM091630
13	19	3.9	582	10	BM082634
14	19	3.9	584	10	BM083068
15	19	3.9	585	10	BI474188
16	19	3.9	617	10	BM160573
17	19	3.9	723	12	AG132034

C 18	19	3.9	852	12	AQ742455
C 19	19	3.9	1179	10	BG114834
C 20	19	3.9	1295	10	BM006536
C 21	18	3.7	163	9	AI081506
C 22	18	3.7	200	9	AI205976
C 23	18	3.7	200	9	AI205977
C 24	18	3.7	200	9	AI820806
C 25	18	3.7	214	9	BB416046
C 26	18	3.7	276	9	BB480551
C 27	18	3.7	284	9	BB198207
C 28	18	3.7	286	9	AA787969
C 29	18	3.7	359	10	BE949717
C 30	18	3.7	373	10	H81695
C 31	18	3.7	379	10	BE952725
C 32	18	3.7	407	12	AQ339510
C 33	18	3.7	408	12	AZ119827
C 34	18	3.7	417	10	BE672478
C 35	18	3.7	424	10	BF223697
C 36	18	3.7	436	10	N81359
C 37	18	3.7	463	10	BE335854
C 38	18	3.7	466	9	AM123846
C 39	18	3.7	468	9	AM221110
C 40	18	3.7	469	10	C87441
C 41	18	3.7	479	10	BG202620
C 42	18	3.7	484	10	BG067771
C 43	18	3.7	485	9	AV847706
C 44	18	3.7	491	9	AM449278
C 45	18	3.7	492	9	AI465386

#### ALIGNMENTS

RESULT 1  
BB517742  
LOCUS  
DEFINITION  
BB517742 RIKEN full-length enriched, 16 days neonate heart Mus  
musculus cDNA clone DB30025004 3', mRNA sequence.

ACCESSION  
BB517742 GI:16443203

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

BB517742 661 bp mRNA linear EST 25-OCT-2001  
BB517742 RIKEN full-length enriched, 16 days neonate heart Mus  
musculus cDNA clone DB30025004 3', mRNA sequence.  
BB517742.2 GI:16443203  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 661)  
Arakawa, T., Carrincci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda  
M., Koyu, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki  
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
Muramatsu, M., and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
On Jul 28, 2000 this sequence version replaced gi:9569200.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carrincci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Matsuhiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Herrmann, R.G., Holtan, T., Jacquemin, J.M., Jia, J., Joudrier, P., Langridge, P., Izzo, G.R., Lin, J.J., McGuire, P., Pecchioni, N., Quesnel, C., Schuch, W., Selvaraj, G., Ogihara, Y., Sorrells, M., Warburton, M., and Wenzel, G., International Triticaceae EST Cooperative (ITREC): Production of International Triticaceae EST Cooperative (ITREC): Production of Expressed Sequence Tags for Species of the Triticaceae Unpublished (2000)  
 Contact: Joudrier, P  
 INRA, Unite de Biochimie et Biologie Molculaire des Cereales  
 2, place VIALA, 34060 Montpellier cedex 01 FRANCE  
 Tel: 33 4 99 61 23 84  
 Fax: 33 4 99 61 23 48  
 Email: joudrier@enscm.inra.fr  
 international.triticaceae.estr.cooperative (ITREC)  
<http://wheat.pw.usda.gov/genome>.  
 Location/Qualifiers  
 1. .395  
 /organism="Triticum turgidum subsp. durum"  
 /culture="wheat"

BASE COUNT	polylinker site.	primers used.	See p50RT1
98 a	111 c	0.3-2.0 kbp average	Insert size."
ORIGIN	114 g	72 t	

	Query Match	4.18;	Score 20;	DB 10;	Length 395;
	Best Local Similarity	100.0%;	Pred. No. 6.9;		
	Matches 20; Conservative	0;	Mismatches	0;	Gaps 0;
OY	116 GCGGTGTTCTGGCCACGCTG	135			
Dd	48 GCGGTGTTCTGGCCACGCTG	29			

transcriptase and subsequently thermo-activated reverse cap-trapper second strand enriched for full-length by primer adapter sequence [5].  
GAGAGAGAGATTCCTCAAGTTAATAATTAAATCCCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified Bluescript KS(+) after blunt excision from Lambda FLX I.\*

BASE COUNT		175	a	162	c	171	g	133	t
ORIGIN									

```

Query Match          4.38; Score 21; DB 9; Length 661;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0.

```

QY	171	TGCTGCTGCTGTTTGGGTGG	191
Db	505	TGCTGCTGCTGTTTGGGTGG	525

RESULT 2	
BE428593/C	
LOCUS	
DEFINITION	BE428593
ACCESSION	MTD008.G05P990617
VERSION	tturgidum subsp. durum MTD Durum Wheat Root Library EST 26-JUL-2000
KEYWORDS	BE428593
SOURCE	BE428593.1 GI:9426436
ORGANISM	durum wheat.
	triticum durum

REFERENCE  
1 (bases 1 to 395)  
Anderson, O.A., Annele P. ...  
... Poaceae; Poaceae; Poaceae

Brave, T., Close, T., Cloutier

```

FEATURES
  source
    1. 673
      location/Qualifiers
      /organism="Medicago truncatula"
      /db_xref="taxon:3880"

```

```

/organism="Medicago truncatula"
/db_xref="taxon:3880"

```

/clone="NF026G01EC"  
/tissue\_type="Elicited cell culture"  
/dev\_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"  
/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

BASE COUNT 175 a 151 c 127 g 220 t

ORIGIN

Query Match 4.1%; Score 20; DB 10; Length 673;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 224 ATCAGCAGATAAATTAAG 243  
Db 343 ATCAGCAGATAAATTAAG 324

RESULT 4  
AZ731780/c 689 bp DNA linear GSS 25-JAN-2001  
LOCUS RPCI-24-128C18.TJ RPCI-24 Mus musculus genomic clone RPCI-24-128C18  
DEFINITION  
; DNA sequence.  
ACCESSION AZ731780  
VERSION AZ731780.1 GI:12494163  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 689)  
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akirel,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-24 unpublished (1999)  
Contact: Shiyang Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac\_end\_intro.html  
Plate: 128 row: C column: 18  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1..689  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-128C18"  
/clone\_lib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: PTARBAC1; Site\_1: BamHI; Site\_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the PTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 166 a 163 c 155 g 205 t

ORIGIN

Query Match 4.1%; Score 20; DB 12; Length 689;

Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 409 GCAGGGCAGACGACGAT 428  
Db 543 GCAGGGCAGACGACGAT 524

RESULT 5  
CNS06G1H/c 988 bp DNA linear GSS 17-JUN-2001  
LOCUS T3 end of clone AS0AA005E09 of library AS0AA from strain CLIB 533  
DEFINITION  
of Saccharomyces bayanus, genomic survey sequence.  
ACCESSION AL397851  
VERSION AL397851.1 GI:12150837  
KEYWORDS GSS.  
SOURCE Saccharomyces bayanus.  
ORGANISM Saccharomyces bayanus.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
1 (bases 1 to 988)  
Bon,E., Neuvéglise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.  
Genomic exploration of the hemiascomycetous yeasts: 5.  
Saccharomyces bayanus var. uvarum  
FEMS Lett. 487 (1), 37-41 (2000)

JOURNAL MEDLINE 20584715  
REFERENCE 2 (bases 1 to 988)  
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogiropoulos,O., Potter,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
FEMS Lett. 487 (1), 3-12 (2000)

JOURNAL MEDLINE 20584711  
REFERENCE 3 (bases 1 to 988)  
Genoscope.  
Direct Submision  
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqrefgenoscope.cns.fr - Web: www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES  
source  
1..988  
Location/Qualifiers  
/organism="Saccharomyces bayanus"  
/strain="CLIB 533"  
/variety="uvarum"  
/db\_xref="taxon:4931"  
/clone="AS0AA005E09"  
/clone\_lib="AS0AA"  
/note="end: T3"

misc-feature  
/note="similar to Saccharomyces cerevisiae ORF YER075c [ PP3 : protein tyrosine phosphatase ]"

BASE COUNT 313 a 253 c 162 g 255 t

ORIGIN

Query Match 4.1%; Score 20; DB 12; Length 988;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 251 AGAGTTTGGAAAGCTAAAA 270  
 Db 622 AGAGTTTGGAAAGCTAAAA 603

RESULT 6  
 LOCUS AU003546  
 DEFINITION AU003546 Selected chromosome 21 cDNA library Homo sapiens CDNA  
 ACCESSION AU003546  
 VERSION AU003546  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE An integrated transcript map for the whole human chromosome 21  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Yasuo M.-L.  
 FEATURES  
 SOURCE Max Planck Institut fuer Molekulare Genetik  
 Location/Qualifiers  
 1..376  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="21q"  
 /clone="Mpi112-9K8"  
 /clone\_lib="Selected chromosome 21 cDNA library"

BASE COUNT 106 a 81 c 81 g 103 t 5 others  
 ORIGIN

Query Match  
 Best Local Similarity 3.9%; Score 19; DB 9; Length 376;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 406 ATTGCAGGCGAGAGCANG 424  
 Db 37 ATTGCAGGCGAGAGCANG 55

RESULT 7  
 LOCUS T59966  
 DEFINITION y67608.r1 Stratagene ovary (#937217) Homo sapiens linear EST 09-FEB-1995  
 IMAGE:76262 5' similar to SP:ICPD\_MOUSE P80315 T-COMPLEX PROTEIN 1,  
 ACCESSION T59966  
 VERSION T59966.1 GI:661803  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE Chissee, S., Dietrich, N., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Hillier, L., Lennon, G., Dubuque, T., Favello, A., Gish, W., Hawkins,  
 'M., Muller, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore  
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,  
 Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.,  
 Generation and analysis of 280,000 human expressed sequence tags  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 MEDLINE 97044478  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

FEATURES  
 source  
 Location/Qualifiers  
 1..380  
 /organism="Homo sapiens"  
 /db\_xref="GDB:505991"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:76262"  
 /clone\_lib="Stratagene ovary (#937217)"  
 /sex="Female"  
 /dev\_stage="49 year old"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: ovary; Vector: Bluescript SK; Site: 1: EcoRI;  
 Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dT.  
 Total ovary tissue, normal, caucasian. Average insert  
 size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
 GAATTCGCGACGAG 3' -3' adaptor sequence: 5'  
 CTCGAGTTTCTTTTCTTTT 3'."  
 High quality sequence stop: 257.  
 Seq primer: M13Rp1  
 Insert length: 1282 Std Error: 0.00  
 Consortium (info@image.llnl.gov) through LNL; contact the IMAGE  
 Consortium (info@image.llnl.gov) for further information.  
 Email: est@watson.wustl.edu  
 Insert Size: 1282  
 High quality sequence stops: 257 Source: IMAGE Consortium, LNL. This  
 clone is available royalty-free through LNL; contact the IMAGE  
 Consortium (info@image.llnl.gov) for further information.

BASE COUNT 104 a 68 c 93 g 111 t 4 others  
 ORIGIN

Query Match  
 Best Local Similarity 3.9%; Score 19; DB 10; Length 380;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 AACCCAGTTTGGAAAGG 265  
 Db 283 AACCCAGTTTGGAAAGG 301

RESULT 8  
 LOCUS AW151916/c  
 DEFINITION x172906.x1 NCI-CGAP-Gas4 Homo sapiens CDNA clone IMAGE:2623642 3',  
 ACCESSION AW151916  
 VERSION AW151916.1 GI:6199901  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE NCI-CGAP htp://www.ncbi.nlm.nih.gov/ncicgap.  
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/bbtp/image/image.html

FEATURES  
 source  
 Location/Qualifiers  
 1..436  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2623642"  
 /clone\_lib="NCI-CGAP-Gas4"

Possible reversed clone: polyT not found  
 Seq primer: -40UP from gibco  
 High quality sequence stop: 412.  
 Location/Qualifiers  
 1..436  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2623642"  
 /clone\_lib="NCI-CGAP-Gas4"

```

/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pCMV-SPORT6; Site_1: Salt,
Site_2: Notti; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT      136 a      98 c      86 g      116 t
ORIGIN

Query Match      3.9%; Score 19; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      299 CAGTACAGCAGCATAGTTA 317
Db      256 CAGATACAGCAGCATAGTTA 238

RESULT 9
AZ952777      461 bp      DNA      linear      GSS 27-APR-2001
LOCUS      2M0217C15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION      clone UUGC2M0217C15 R, DNA sequence.
ACCESSION      AZ952777
VERSION      AZ952777.1 GI:13824004
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
1 (bases 1 to 461)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weis,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0217 row: C column: 15
Seq primer: CACACAGAAACAGCTGTGACC
Class: plasmid ends
High quality sequence stop: 461.
Location/Qualifiers
1. 461
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0217C15"
/clone_1lb="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

```

```

of pMD42 (g14732114|p|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      123 a      91 c      121 g      126 t
ORIGIN

Query Match      3.9%; Score 19; DB 12; Length 461;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      170 TTGGTGGTCTGTTTGGG 188
Db      413 TTGGTGGTCTGTTTGGG 431

RESULT 10
A1965711      545 bp      mRNA      linear      EST 30-NOV-2001
LOCUS      sc77604.y1 Gm-c1018 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION      Gm-c1018-919 5' similar to TR:080558 080558 722J18.21 PROTEIN. ;,
mRNA sequence.
ACCESSION      A1965711
VERSION      A1965711.1 GI:5760420
KEYWORDS      EST.
SOURCE      soybean.
ORGANISM      Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 545)
Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,D., Corvelli,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers
,T., Peterson,B., Swaller,T., Gibbons,M., Pepe,D., Harvey,N., Schuk
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
possible reversed clone: similarity on wrong strand This clone is
available through: Resgen, Invitrogen Corp. 2130 South Memorial
Parkway Huntsville, AL 35801 For further information call: (800
)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 421.
Location/Qualifiers
1. 545
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1018-919"
/clone_1lb="Gm-c1018"
/tissue_type="leaves of greenhouse grown plants"
/dev_stage="2-3 weeks old"
/lab_host="DH10B (Gibco BRL)"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from 2-3 week old greenhouse grown plants. The cDNA library
was prepared using the Stratagene pluscript II XR
library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dt) sequence with a XhoI restriction site. EcoRI adaptors
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally

```

cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg.

BASE COUNT 147 a 115 c 120 g 163 t

Query Match 3.9%; Score 19; DB 9; Length 545;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 GTTTGGGTGATTAATG 199  
|||||

DB 306 GTTTGGGTGATTAATG 324

RESULT 11  
AZ887587  
LOCUS RPCI-24-166M15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-166M15  
DEFINITION ' DNA sequence.  
ACCESSION AZ887587  
VERSION AZ887587.1 GI:13206532  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 550)

TITLE Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akimov, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.  
JOURNAL Mouse BAC End Sequences from Library RPCI-24  
COMMENT Unpublished (1999)  
Other GSSs: RPCI-24-166M15.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end plates: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
Plate: 166 row: M column: 15  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
Source Location/Qualifiers

1..550  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-166M15"  
/clone\_1ib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pPARBAC1; Site: 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pPARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 126 a 155 c 123 g 146 t

Query Match 3.9%; Score 19; DB 12; Length 550;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 170 TTGGTGTCTCTTTTGGG 188

DB 213 TTGGTGTCTCTTTTGGG 231  
|||||

RESULT 12  
BM091630  
LOCUS BM091630 561 bp mRNA linear EST 29-NOV-2001  
DEFINITION sag99e06.y1 Gm-cl086 glycine max cDNA clone GENOME SYSTEMS CLONE  
ID: Gm-cl086-203 5' similar to TR:09T0C8 09T0C8 PUTATIVE PHOSPHORIBOSYLAMINTRANSLATE TRANSFERASE. ; mRNA sequence.  
ACCESSION BM091630  
VERSION BM091630.1 GI:17020596  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine. 1 (bases 1 to 561)

Shoemaker, R., Keim, P., Vodkin, L., Erpelidg, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Rheising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ralter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available through: Resgen, Invitrogen Corp, 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: [cvu@resgen.com](mailto:cvu@resgen.com)  
High quality sequence stop: 421.

FEATURES  
Source Location/Qualifiers

1..561  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl086-203"  
/clone\_1ib="Gm-cl086"  
/tissue\_type="young seeds (Williams 82)"  
/lab\_host="DH10B"  
/note="Vector: pBluescript II SK+; Site: 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from very young seeds (less than 20mgs). The library was prepared using the Stratagene pBluescript II SK (+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an Xho I restriction site. Eco RI adaptors were ligated to the blunt-ended cDNA fragments followed by Xho I digestion. The cDNA insert is protected from Xho I digestion via methylation during first strand cDNA synthesis. The cDNA fragments were directionally cloned into the Eco RI-Xho I restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cell. The library was constructed by Anu Khanna (Ulla Vodkin lab, University of Illinois)."

BASE COUNT 126 a 136 c 134 g 165 t

Query Match 3.9%; Score 19; DB 10; Length 561;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 GTTTGGGTGATTAATG 199  
|||||

DB 4 GTTTGGGTGATTAATG 22



```

RESULT 13
BM082634/c 582 bp mRNA linear EST 14-NOV-2001
LOCUS f423j09.y1 Campbell zebrafish ovary danio rerio cDNA clone 5306345
DEFINITION 5' similar to TR:Q9Y6R2 Q9Y6R2 EMBRYONIC LUNG PROTEIN. ; mRNA
sequence.
ACCESSION BM082634
VERSION BM082634.1 GI:16929564
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 582)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
Mashu zebrafish EST Project 1998
Unpublished (1998)
TITLE Mashu zebrafish EST Project 1998
JOURNAL
COMMENT
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CDNA library constructed by library constructed by Invitrogen and
donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: the I.M.A.G.E. Consortium/LNL, send email to:
info@image.llnl.gov
High quality sequence stop: 440.
FEATURES
Location/Qualifiers
source
1..582
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5306345"
/clone_lib="Campbell zebrafish ovary"
/sex="female"
/dev_stage="4-5 months, 1 year and 2 years"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
NOT1; Site_2: SalI; Whole ovaries collected from zebrafish
aged 4-5 months, 1 year and 2 years. Oligo-dT primed,
directionally cloned. Average insert size 2 kb. Library
constructed by Invitrogen and donated by R. Campbell
(Marine Biology Laboratory, Woods Hole, MA)."
```

```

SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 584)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
Mashu zebrafish EST Project 1998
Unpublished (1998)
TITLE Mashu zebrafish EST Project 1998
JOURNAL
COMMENT
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CDNA library constructed by library constructed by Invitrogen and
donated by R. Campbell (Marine Biology Laboratory, Woods Hole, MA).
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: the I.M.A.G.E. Consortium/LNL, send email to:
info@image.llnl.gov
High quality sequence stop: 442.
FEATURES
Location/Qualifiers
source
1..584
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5306748"
/clone_lib="Campbell zebrafish ovary"
/sex="female"
/dev_stage="4-5 months, 1 year and 2 years"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary(pooled); Vector: pCMV-SPORT6; Site_1:
NOT1; Site_2: SalI; Whole ovaries collected from zebrafish
aged 4-5 months, 1 year and 2 years. Oligo-dT primed,
directionally cloned. Average insert size 2 kb. Library
constructed by Invitrogen and donated by R. Campbell
(Marine Biology Laboratory, Woods Hole, MA)."
```

```

BASE COUNT 166 a 145 c 137 g 133 t
ORIGIN
Query Match 3.9%; Score 19; DB 10; Length 582;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 240 AAACCTAACCAAGCTTGG 258
|||||
Db 207 AAACCTAACCAAGCTTGG 189
```

```

RESULT 15
BI474188/c 585 bp mRNA linear EST 27-AUG-2001
LOCUS fp36b02.y3 zebrafish gridded kidney danio rerio cDNA clone 4745043
DEFINITION 5' similar to TR:Q9Y6R2 Q9Y6R2 EMBRYONIC LUNG PROTEIN. ; mRNA
sequence.
ACCESSION BI474188
VERSION BI474188.1 GI:15303352
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 585)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
```

TITLE  
JOURNAL  
COMMENT

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.  
and Wilson, R.  
Washu zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
CDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:  
Washington University Genome Sequencing Center Clone distribution:  
Genome Systems, St. Louis, Missouri (web address:  
www.genomesystems.com) (email contact: info@genomesystems.com)  
Research Genetics, Huntsville, Alabama (web address: www.resgen.com)  
) (email contact: info@resgen.com) and  
Ressourcenzentrum Primatardatenbank, Berlin, Germany (web address:  
www.rzpd.de)  
High quality sequence stop: 320.

FEATURES  
source

Location/Qualifiers  
1..585  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="4745043"  
/clone\_lib="zebrafish gridded kidney"  
/sex="mixed"  
/tissue\_type="kidney pooled from 300 wild type adults"  
/lab\_host="XLOLR"  
/note="Organ: kidney; Vector: pBK-CMV; Site 1: ECORI;  
Site 2: XhoI; Oligo dt cDNA library constructed from mRNA  
pooled from pooled kidney tissue from 300 adult  
zebrafish."  
BASE COUNT 161 a 140 c 136 g 146 t 2 others  
ORIGIN

Query Match 3.9%; Score 19; DB 10; Length 585;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 240 AAAGCTAAACGAGGTTTG 258  
|||||  
DB 291 AAAGCTAAACGAGGTTTG 273

Search completed: October 27, 2002, 22:36:36  
Job time : 978.679 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:19:49 : Search time 36.0933 Seconds  
(without alignments)  
3307.484 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486  
Sequence: 1 atgaacagagatgttgc.....aagtcattcacacagaaaa 486

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*
  - 3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*
  - 4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/1na/PCTUS.COMB.seq:\*
  - 6: /cgn2\_6/ptodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	3.7	1331	US-08-985-908-23	Sequence 23, Appl
2	18	3.7	3776	US-08-162-809-7	Sequence 7, Appl
3	17	3.5	2165	US-08-669-721-2	Sequence 2, Appl
4	17	3.5	2165	US-09-189-344-2	Sequence 2, Appl
5	17	3.5	2340	US-08-742-877-3	Sequence 3, Appl
6	17	3.5	2775	US-09-053-871A-22	Sequence 22, Appl
7	17	3.5	2802	US-08-742-877-1	Sequence 1, Appl
8	17	3.5	51952	US-08-947-823-1	Sequence 1, Appl
9	17	3.5	4403765	US-09-103-840A-2	Sequence 2, Appl
10	16	3.3	1632	US-07-959-941-1	Sequence 1, Appl
11	16	3.3	1632	US-08-259-924-1	Sequence 1, Appl
12	16	3.3	2251	US-08-991-677-11	Sequence 11, Appl
13	16	3.3	3429	US-08-097-997A-10	Sequence 10, Appl
14	16	3.3	3429	US-08-665-574C-10	Sequence 10, Appl
15	16	3.3	3429	US-08-946-994-10	Sequence 10, Appl
16	16	3.3	4234	US-08-446-038B-1	Sequence 1, Appl
17	16	3.3	4234	US-08-446-010B-1	Sequence 1, Appl
18	16	3.3	4234	US-08-805-445-1	Sequence 1, Appl
19	16	3.3	4234	US-08-064-067D-1	Sequence 1, Appl
20	16	3.3	4234	US-09-066-208-1	Sequence 1, Appl
21	16	3.3	4868	US-08-139-937-12	Sequence 12, Appl
22	16	3.3	4868	PCT-US93-11310-12	Sequence 12, Appl
23	16	3.3	5703	US-08-280-590A-36	Sequence 36, Appl
24	16	3.3	8789	US-08-328-254-5	Sequence 5, Appl
25	16	3.3	10136	US-08-353-700-2	Sequence 2, Appl
26	16	3.3	10136	PCT-US95-16216-2	Sequence 2, Appl
27	16	3.3	24417	US-08-846-762-1	Sequence 1, Appl

C 28	15	3.1	59	4	US-09-037-990B-64	Sequence 64, Appl
C 29	15	3.1	99	1	US-08-248-474-77	Sequence 77, Appl
C 30	15	3.1	99	3	US-08-756-849-77	Sequence 77, Appl
C 31	15	3.1	162	4	US-09-037-990B-36	Sequence 36, Appl
C 32	15	3.1	177	1	US-08-065-845-9	Sequence 9, Appl
C 33	15	3.1	177	1	US-08-429-523-9	Sequence 9, Appl
C 34	15	3.1	177	1	US-08-429-532-9	Sequence 9, Appl
C 35	15	3.1	177	1	US-08-429-532-9	Sequence 9, Appl
C 36	15	3.1	177	1	US-08-429-532-9	Sequence 9, Appl
C 37	15	3.1	177	1	US-08-429-532-9	Sequence 9, Appl
C 38	15	3.1	249	1	US-08-269-136B-9	Sequence 9, Appl
C 39	15	3.1	249	3	US-08-248-474-4	Sequence 4, Appl
C 40	15	3.1	354	2	US-08-756-849-4	Sequence 4, Appl
C 41	15	3.1	379	4	US-08-676-279-38	Sequence 38, Appl
C 42	15	3.1	422	4	US-08-838-151A-58	Sequence 58, Appl
C 43	15	3.1	494	4	US-09-037-990B-12	Sequence 12, Appl
C 44	15	3.1	494	4	US-09-040-984-54	Sequence 54, Appl
C 45	15	3.1	725	3	US-09-123-912-54	Sequence 54, Appl
					US-09-197-801-12	Sequence 12, Appl

## ALIGNMENTS

RESULT 1  
US-08-985-908-23  
Sequence 23, Application US/08985908  
Patent No. 6004773  
GENERAL INFORMATION:  
APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHINARA, AND TSUYOSHI  
TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
CITY: ARLINGTON  
COUNTY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/985,908  
FILING DATE: 05-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-325659  
FILING DATE: 05-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: NORMAN F. OBLON  
REGISTRATION NUMBER: 24,618  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2230  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1331 bases  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORGANISM: Escherichia coli  
STRAIN: JM109  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10..1197  
US-08-985-908-23  
Query Match 3.7%; Score 18; DB 3; Length 1331;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 GCTGTTGGTAAAGTAG 149  
|||||  
Db 252 GCTGTTGGTAAAGTAG 269

## RESULT 2

US-08-162-809-7  
Sequence 7, Application US/08162809  
Patent No. 5457048

GENERAL INFORMATION:  
APPLICANT: Pasquale, Elena B.  
APPLICANT: Sajjadi, Rezaeioun G.  
TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CAMPBELL AND FLORES  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States of America  
ZIP: 92122

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/162.809  
FILING DATE:

CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 9503  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3776 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 290..3208  
US-08-162-809-7

Query Match  
Best Local Similarity 3.7%; Score 18; DB 1; Length 3776;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 GGCATTGGTGGCTGT 182  
|||||  
Db 271 GGCATTGGTGGCTGT 288

## RESULT 3

US-08-669-721-2/c  
Sequence 2, Application US/08669721  
Patent No. 5834236

GENERAL INFORMATION:  
APPLICANT: Lamb et al., Christopher J.  
TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA

COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669.721  
FILING DATE: 27-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ellison, Eldora L.  
REGISTRATION NUMBER: 39,967  
REFERENCE/DOCKET NUMBER: 07251/014001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2165 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..1698  
US-08-669-721-2

Query Match  
Best Local Similarity 3.5%; Score 17; DB 2; Length 2165;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 ACTGGGCTGTTGTCG 116  
|||||  
Db 673 ACTGGGCTGTTGTCG 657

## RESULT 4

US-09-189-344-2/c  
Sequence 2, Application US/09189344  
Patent No. 6191258

GENERAL INFORMATION:  
APPLICANT: Lamb et al., Christopher J.  
TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/189.344  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/669,721  
FILING DATE: 27-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ellison, Eldora L.  
REGISTRATION NUMBER: 39,967  
REFERENCE/DOCKET NUMBER: 07251/014001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2165 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..1698  
US-09-189-344-2

Query Match 3.5%; Score 17; DB 4; Length 2165;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 ACTGGGCTGTTGTCG 116  
|||||  
DB 673 ACTGGGCTGTTGTCG 657

## RESULT 5

US-08-742-877-3/c  
Sequence 3, Application US/08742877  
Patent No. 6046380  
GENERAL INFORMATION:  
APPLICANT: CLARK, Anthony J.  
TITLE OF INVENTION: DNA SEQUENCES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/742,877  
FILING DATE: 01-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9408717.8  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: FLESHNER, RAZ E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 0623.0470001/REF  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2340 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
US-08-742-877-3

Query Match 3.5%; Score 17; DB 3; Length 2340;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 255 TTGGAAGGTAAG 271  
|||||  
DB 1495 TTGGAAGGTAAG 1479

## RESULT 6

US-09-053-871A-22/c  
Sequence 22, Application US/09053871A  
Patent No. 6315995  
GENERAL INFORMATION:  
APPLICANT: Pinsky, David J.  
APPLICANT: Stern, David  
APPLICANT: Rose, Eric  
APPLICANT: Solomon, Robert A.  
APPLICANT: Schmidt, Ann Marie  
TITLE OF INVENTION: METHODS FOR TREATING AN ISCHEMIC DISORDER AND IMPROVING  
FILE REFERENCE: 51917-B  
CURRENT APPLICATION NUMBER: US/09/053,871A  
CURRENT FILING DATE: 1998-04-01  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 22  
LENGTH: 2775  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-09-053-871A-22

Query Match 3.5%; Score 17; DB 4; Length 2775;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 255 TTGGAAGGTAAG 271  
|||||  
DB 1928 TTGGAAGGTAAG 1912

## RESULT 7

US-08-742-877-1/c  
Sequence 1, Application US/08742877  
Patent No. 6046380  
GENERAL INFORMATION:  
APPLICANT: CLARK, Anthony J.  
TITLE OF INVENTION: DNA SEQUENCES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/742,877  
FILING DATE: 01-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9408717.8  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: FLESHNER, RAZ E.  
REGISTRATION NUMBER: 34,331  
REFERENCE/DOCKET NUMBER: 0623.0470001/REF  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2802 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 30..1412  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 30..167  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 168..1412  
US-08-742-877-1

Query Match 3.5%; Score 17; DB 3; Length 2802;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 TTGGGAAAGGTAAAG 271  
DB 1957 TTGGGAAAGGTAAAG 1941

RESULT 8  
US-08-947-823-1  
Sequence 1, Application US/08947823  
Patent No. 6114605  
GENERAL INFORMATION:  
APPLICANT: Williamson, Valerie M.  
APPLICANT: Kaloshian, Isouh  
APPLICANT: Yaghoobi, Jafar  
APPLICANT: Bodeau, John  
APPLICANT: Milligan, Stephen  
TITLE OF INVENTION: Procedures and Materials for Conferring  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/947,823  
FILING DATE: 09-OCT-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/18802  
FILING DATE: 09-OCT-1997  
APPLICATION NUMBER: US 60/028,191  
FILING DATE: 10-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-0702100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51952 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-947-823-1

Query Match 3.5%; Score 17; DB 3; Length 51952;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 TCATTCAACAGAAAA 486  
DB 49389 TCATTCAACAGAAAA 49405

RESULT 9  
US-09-103-840A-2/C  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2 4403765  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2

Query Match 3.5%; Score 17; DB 4; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GCGCGTGTCTGCGCCA 131  
DB 4088325 GCGCGTGTCTGCGCCA 4088309

RESULT 10  
US-07-959-941-1  
Sequence 1, Application US/07959941  
Patent No. 5364781  
GENERAL INFORMATION:  
APPLICANT: HUTCHINSON, Charles R.  
APPLICANT: MADDDRI, Krishna M.  
APPLICANT: TORI, Francesca  
APPLICANT: COLOMBO, Anna L.  
TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/959,941  
FILING DATE: 19921009  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 793,873  
FILING DATE: 18-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Chin, Monica F.

```

:      REGISTRATION NUMBER: P-36,105
:      REFERENCE/DOCKET NUMBER: 1615-1816CIP
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: (202)638-5000
:      TELEFAX: (202)638-4810
:      INFORMATION FOR SEQ ID NO: 1:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 1632 base pairs
:      TYPE: NUCLEIC ACID
:      STRANDEDNESS: double
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)
:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: 204..1271
:      US-07-959-941-1
:
:      Query Match          3.3%; Score 16; DB 1; Length 1632;
:      Best Local Similarity 100.0%; Pred. No. 47;
:      Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
:      Oy      439 GCATGCCGCAACCGG 454
:      Db      1 GCATGCCGCAACCGG 16
:
:      RESULT 11
:      US-08-259-924-1
:      Sequence 1, Application US/08259924
:      Patent No. 5563064
:      GENERAL INFORMATION:
:      APPLICANT: HUTCHINSON, Charles R.
:      APPLICANT: MADURIT, Krishna M.
:      APPLICANT: TORRI, Francesca
:      APPLICANT: COLOMBO, Anna L.
:      TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN
:      NUMBER OF SEQUENCES: 6
:      CORRESPONDENCE ADDRESS:
:      ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
:      STREET: 655 Fifteenth Street N.W. Suite 330
:      CITY: Washington
:      STATE: D.C.
:      COUNTRY: U.S.A.
:      ZIP: 20005-5701
:      COMPUTER READABLE FORM:
:      MEDIUM TYPE: Floppy disk
:      COMPUTER: IBM PC compatible
:      OPERATING SYSTEM: PC-DOS/MS-DOS
:      SOFTWARE: Patentin Release #1.0, Version #1.25
:      CURRENT APPLICATION DATA:
:      APPLICATION NUMBER: US/08/259,924
:      FILING DATE:
:      CLASSIFICATION: 435
:      PRIOR APPLICATION DATA:
:      APPLICATION NUMBER: US 793,873
:      FILING DATE: 18-NOV-1991
:      PRIOR APPLICATION DATA:
:      APPLICATION NUMBER: US 07/959,941
:      FILING DATE: 09-OCT-1992
:      ATTORNEY/AGENT INFORMATION:
:      NAME: Kitz, Monica C
:      REGISTRATION NUMBER: 36,105
:      REFERENCE/DOCKET NUMBER: 1615-4003
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: (202)638-5000
:      TELEFAX: (202)638-4810
:      INFORMATION FOR SEQ ID NO: 1:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 1632 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: double
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)

```

```

:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: 204..1271
:      US-08-259-924-1
:
:      Query Match          3.3%; Score 16; DB 1; Length 1632;
:      Best Local Similarity 100.0%; Pred. No. 47;
:      Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
:      Oy      439 GCATGCCGCAACCGG 454
:      Db      1 GCATGCCGCAACCGG 16
:
:      RESULT 12
:      US-08-991-677-11
:      Sequence 11, Application US/08991677A
:      Patent No. 6252135
:      GENERAL INFORMATION:
:      APPLICANT: Chiang, Vincent L
:      APPLICANT: Carraway, Daniel T
:      APPLICANT: Smeltzer, Richard H
:      TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
:      FILE REFERENCE: 50617
:      CURRENT APPLICATION NUMBER: US/08/991,677A
:      CURRENT FILING DATE: 1997-12-16
:      EARLIER APPLICATION NUMBER: US 60/033,381
:      EARLIER FILING DATE: 1996-12-16
:      NUMBER OF SEQ ID NOS: 11
:      SOFTWARE: Patentin Ver. 2.0
:      SEQ ID NO 11
:      LENGTH: 2251
:      TYPE: DNA
:      ORGANISM: Pinus taeda
:      US-08-991-677-11
:
:      Query Match          3.3%; Score 16; DB 4; Length 2251;
:      Best Local Similarity 100.0%; Pred. No. 48;
:      Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
:      Oy      234 TAAATTAAGCTAAGC 249
:      Db      310 TAAATTAAGCTAAGC 325
:
:      RESULT 13
:      US-08-097-997A-10/C
:      Sequence 10, Application US/08097997A
:      Patent No. 5728536
:      GENERAL INFORMATION:
:      APPLICANT: Ihle, James N.
:      APPLICANT: Silvenoinen, Olli
:      APPLICANT: Witthuhn, Bruce A.
:      APPLICANT: Ouellet, Frederick W.
:      TITLE OF INVENTION: Jak kinases and Regulation of Cytokine Signal
:      Transduction
:      NUMBER OF SEQUENCES: 14
:      CORRESPONDENCE ADDRESS:
:      ADDRESSEE: Sterne, Kessler, Goldstein & Fox
:      STREET: 1100 New York Avenue, Suite 600
:      CITY: Washington
:      STATE: D.C.
:      COUNTRY: U.S.A.
:      ZIP: 20005-3934
:      COMPUTER READABLE FORM:
:      MEDIUM TYPE: Floppy disk
:      COMPUTER: IBM PC compatible
:      OPERATING SYSTEM: PC-DOS/MS-DOS
:      SOFTWARE: Patentin Release #1.0, Version #1.25
:      CURRENT APPLICATION DATA:
:      APPLICATION NUMBER: US/08/097,997A
:      FILING DATE: 29-JULY-1994
:      CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:  
NAME: Fox, Samuel L.  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 0656.0370000/SLF/GKT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2540  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3429 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3426  
US-08-097-997A-10

Query Match  
Best Local Similarity 3.3%; Score 16; DB 1; Length 3429;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 125 CTGGCCAGCTGTTGG 140  
Db 2063 CTGGCCAGCTGTTGG 2048  
|||||

RESULT 14  
US-08-665-574C-10/C  
Sequence 10, Application US/08665574C  
Patent No. 6136595  
GENERAL INFORMATION:  
APPLICANT: Ihle, James N.  
APPLICANT: Silvenoinen, Olli  
APPLICANT: Wilthuhn, Bruce A.  
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,574C  
FILING DATE: 18-JUN-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/282,012  
FILING DATE: 29-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/097,997  
FILING DATE: 29-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/118,968  
FILING DATE: 09-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0656.0370002/SLF/LBB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3429 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3426  
US-08-665-574C-10

Query Match  
Best Local Similarity 3.3%; Score 16; DB 3; Length 3429;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 125 CTGGCCAGCTGTTGG 140  
Db 2063 CTGGCCAGCTGTTGG 2048  
|||||

RESULT 15  
US-08-946-994-10/C  
Sequence 10, Application US/08946994  
Patent No. 6210654  
GENERAL INFORMATION:  
APPLICANT: Ihle, James N.  
APPLICANT: Silvenoinen, Olli  
APPLICANT: Wilthuhn, Bruce A.  
APPLICANT: Quelle, Frederick W.  
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,994  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,574  
FILING DATE: 18-JUN-1996  
APPLICATION NUMBER: 08/282,012  
FILING DATE: 29-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/097,997  
FILING DATE: 29-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/118,968  
FILING DATE: 09-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, Samuel L.  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 0656.0370002/SLF/GKT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3429 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3426  
US-08-946-994-10



Query Match 3.3%; Score 16; DB 4; Length 3429;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125 CTGCCCAGCTGTTGG 140  
|||||  
Db 2063 CTGCCCAGCTGTTGG 2048

Search completed: October 27, 2002, 22:53:59  
Job time : 942.093 secs

**THIS PAGE BLANK (USPTO)**



PA (KAYM) KAY W W.  
PA (BURI) BURIAN J.  
XX (KUZK) KUZK M A.

PI Kay WM, Burian J, Kuzyk MA;

DR MPI: 2001-316844/34.  
DR P-PSDB: AAB81126.

PT Method for protecting polkiothermic fish against salmonid rickettsial  
PT septicemia and other rickettsial diseases comprises administering a  
XX vaccine containing the Ospa protein of Piscirickettsia salmonis  
PS Disclosure; Fig 2B; 35pp; English.

CC This invention relates to a method for the protection against infection  
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a  
CC Ospa in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkiothermic fish, against the bacterial pathogen  
CC P. salmonis. The method is also useful for protecting against salmonid  
CC rickettsial septicemia (SRS) and other rickettsial diseases.  
CC sequence represents DNA encoding the P. salmonis Ospa protein. The present  
CC protein with an N-terminal fusion partner is used in a vaccine to create  
CC an anti-Ospa antibody response.

Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other:

Query Match  
Best Local Similarity 100.0%; Score 486; DB 22; Length 486;  
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGAGATGTTGCAAGTAGTACTAATTAATATCATGCTGTTTACTGGC 60  
DB 1 ATGACAGAGATGTTGCAAGTAGTACTAATTAATATCATGCTGTTTACTGGC 60  
QY 61 TGTGCCAGAACTTACTGCTCAAGAAAGTCGAGCTGGGCGCTGTTTGGCGGT 120  
DB 61 TGTGCCAGAACTTACTGCTCAAGAAAGTCGAGCTGGGCGCTGTTTGGCGGT 120  
QY 121 GTTCTGCGCCAGCTGTTTGTAAAGTAGTGTGCAAGTGGCCATTGGTGGTCT 180  
DB 121 GTTCTGCGCCAGCTGTTTGTAAAGTAGTGTGCAAGTGGCCATTGGTGGTCT 180  
QY 181 GTTTGGGTGATTAATGTTCTAAATCGCTCAATCGATGATGAGAGATAATA 240  
DB 181 GTTTGGGTGATTAATGTTCTAAATCGCTCAATCGATGATGAGAGATAATA 240  
QY 241 AAGCTAAACCAAGTTGGAAGAGTAAAGCAGGCGAATGACAGCTTGGCGTAATCA 300  
DB 241 AAGCTAAACCAAGTTGGAAGAGTAAAGCAGGCGAATGACAGCTTGGCGTAATCA 300  
QY 301 GATACAGGCATAGTAAAGTGTGAGCAGTGCCTACTTACCAAGCTTAACATAACCA 360  
DB 301 GATACAGGCATAGTAAAGTGTGAGCAGTGCCTACTTACCAAGCTTAACATAACCA 360  
QY 361 GAGGCTGCCAGCAATATTGTGAGCAATTTCAAGCAAAAGGAGATGAGAGCGCAAG 420  
DB 361 GAGGCTGCCAGCAATATTGTGAGCAATTTCAAGCAAAAGGAGATGAGAGCGCAAG 420  
QY 421 CAAGAGATTTAGGCAACGTCATGCCGCAACCGATGCTGTTGGCAATTTCAACA 480  
DB 421 CAAGAGATTTAGGCAACGTCATGCCGCAACCGATGCTGTTGGCAATTTCAACA 480  
QY 481 GAAAAA 486  
DB 481 GAAAAA 486

RESULT 2  
AAH79040  
ID AAH79040 standard; cDNA; 489 BP.

XX AAH79040;

DT 15-JAN-2002 (first entry)

DE Piscirickettsia salmonis polynucleotide p10.6.

KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;  
KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;  
KW ATCC VR-1361; SS.

OS Piscirickettsia salmonis.

Key Location/Qualifiers  
FT CDS 1..489  
FT /tag a

W0200168865-A2.

20-SEP-2001.

12-MAR-2001; 2001WO-GB01055.

11-MAR-2000; 2000GB-0005838.

01-JUL-2000; 2000GB-0016080.

29-JUL-2000; 2000GB-0018599.

(AQUA-) AQUA HEALTH EURO LTD.

PI Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;  
PI Burzio L;  
DR MPI: 2001-639050/73.  
DR P-PSDB; AAG78025.

PT New nucleic acids encoding an amino acid sequence homologous to the  
PT surface antigen present on piscirickettsia salmonis are useful to  
PT protect fish against piscirickettsiosis -

PS Claim 3; Fig 4; 25pp; English.

CC The invention relates to nucleic acid sequences and the encoded protein  
CC of a least part of the surface antigen present on piscirickettsia  
CC salmonis for production of a vaccine with antibacterial activity to  
CC protect fish against P. salmonis which causes piscirickettsiosis  
CC known as salmonid rickettsial septicemia.

Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other:

Query Match  
Best Local Similarity 100.0%; Score 486; DB 22; Length 489;  
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGAGATGTTGCAAGTAGTACTAATTAATATCATGCTGTTTACTGGC 60  
DB 1 ATGACAGAGATGTTGCAAGTAGTACTAATTAATATCATGCTGTTTACTGGC 60  
QY 61 TGTGCCAGAACTTACTGCTCAAGAAAGTCGAGCTGGGCGCTGTTTGGCGGT 120  
DB 61 TGTGCCAGAACTTACTGCTCAAGAAAGTCGAGCTGGGCGCTGTTTGGCGGT 120  
QY 121 GTTCTGCGCCAGCTGTTTGTAAAGTAGTGTGCAAGTGGCCATTGGTGGTCT 180  
DB 121 GTTCTGCGCCAGCTGTTTGTAAAGTAGTGTGCAAGTGGCCATTGGTGGTCT 180  
QY 181 GTTTGGGTGATTAATGTTCTAAATCGCTCAATCGATGATGAGAGATAATA 240  
DB 181 GTTTGGGTGATTAATGTTCTAAATCGCTCAATCGATGATGAGAGATAATA 240  
QY 241 AAGCTAAACCAAGTTGGAAGAGTAAAGCAGGCGAATGACAGCTTGGCGTAATCA 300  
DB 241 AAGCTAAACCAAGTTGGAAGAGTAAAGCAGGCGAATGACAGCTTGGCGTAATCA 300

```

QY 301 GATACAGCATAGTATAGTGTGAGCCAGTGGCTACTTACAGCGTTACATTAAGCAA 360
    |||||||
Db 301 GATACAGCATAGTATAGTGTGAGCCAGTGGCTACTTACAGCGTTACATTAAGCAA 360
QY 361 GAGCGTCGCCAGCAATATTTGTCGAAATTTTCAGCAAAAAGCGATGATTCAGGCGAGAG 420
    |||||||
Db 361 GAGCGTCGCCAGCAATATTTGTCGAAATTTTCAGCAAAAAGCGATGATTCAGGCGAGAG 420
QY 421 CAAGAGATTTACGGCACTGATGCGCGCAACCGATGCTGTGGCAAGTCATTTCACAA 480
    |||||||
Db 421 CAAGAGATTTACGGCACTGATGCGCGCAACCGATGCTGTGGCAAGTCATTTCACAA 480
QY 481 GAAAAA 486
    |||||
Db 481 GAAAAA 486

RESULT 3
AAF86249
ID AAF86249 standard; DNA: 33 BP.
AC AAF86249;
XX
XX 11-JUL-2001 (first entry)
DE
XX Forward Ospa PCR primer used in construction of pTYB1-17KD.
XX
XX Polkillothermic fish: Piscirickettsia salmonis: rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer: ss.
XX
XX Piscirickettsia salmonis.
OS Synthetic.
XX
XX CA2281913-A1.
PN
XX 17-MAR-2001.
PD
XX 17-SEP-1999; 99CA-2281913.
PF
XX 17-SEP-1999; 99CA-2281913.
PR
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
XX (KUZV/) KUZIK M A.
XX
XX Kay WW, Burian J, Kuzik MA;
PI
XX WPI: 2001-316844/34.
DR
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 5; Page 21; 35pp: English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used to amplify the P. salmonis Ospa DNA
CC sequence. The primer is used in the construction of pTYB1-17kd, which is
CC used to express optimised Ospa antigen with a C-terminal fusion partner.
CC The Ospa fusion protein is used in a vaccine to create an anti-Ospa
CC antibody response.
XX
XX Sequence 33 BP; 13 A; 3 C; 11 G; 6 T; 0 other;
SO

```

```

Query Match 4.7%; Score 23; DB 22; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACAGAGAGGATTTGCAAGG 23
    |||||||
Db 11 ATGACAGAGAGATTTGCAAGG 33

```

```

RESULT 4
AAF86250/c
ID AAF86250 standard; DNA: 40 BP.
AC AAF86250;
XX
XX 11-JUL-2001 (first entry)
DE
XX Reverse Ospa PCR primer used in construction of pTYB1-17KD.
XX
XX Polkillothermic fish: Piscirickettsia salmonis: rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; 17E2; fusion construct; PCR primer: ss.
XX
XX Piscirickettsia salmonis.
OS Synthetic.
XX
XX CA2281913-A1.
PN
XX 17-MAR-2001.
PD
XX 17-SEP-1999; 99CA-2281913.
PF
XX 17-SEP-1999; 99CA-2281913.
PR
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
XX (KUZV/) KUZIK M A.
XX
XX Kay WW, Burian J, Kuzik MA;
PI
XX WPI: 2001-316844/34.
DR
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 5; Page 21; 35pp: English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used to amplify the P. salmonis Ospa DNA
CC sequence. The primer is used in the construction of pTYB1-17kd, which is
CC used to express optimised Ospa antigen with a C-terminal fusion partner.
CC The Ospa fusion protein is used in a vaccine to create an anti-Ospa
CC antibody response.
XX
XX Sequence 40 BP; 8 A; 10 C; 7 G; 15 T; 0 other;
SO
Query Match 4.5%; Score 22; DB 22; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 465 GCAAGTCATTTCAACAGAAAAA 486
    |||||||
Db 40 GCAAGTCATTTCAACAGAAAAA 19

```

```

RESULT 5
ID AAA96363 standard; DNA; 50000 BP.
XX
XX AAA96363;
AC
XX 08-FEB-2001 (first entry)
XX
DE Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
XX
XX Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
KW Hashimoto's disease; coeliac disease; ss.
XX
XX Homo sapiens.
OS
FH Key
FH Location/Qualifiers
FT 5722..5746
FT /tag= a
FT /note= "sara41/42 microsatellite repeat"
FT 6530..6597
FT /tag= b
FT /note= "sara43/44 microsatellite repeat"
FT 19911..19956
FT /tag= c
FT /note= "PW210/211 microsatellite repeat"
FT 23904..23957
FT /tag= d
FT /note= "sara45/46 microsatellite repeat"
FT 27689..27780
FT /tag= e
FT /note= "sara17/18 microsatellite repeat"
FT 30766..30801
FT /tag= f
FT /note= "sara19/20 microsatellite repeat"
XX
XX W0200056856-A2.
XX
XX 28-SEP-2000.
XX
XX 24-MAR-2000; 2000MO-US07938.
XX
XX 25-MAR-1999; 99US-0126215.
XX
XX (GEM) GENETICS INST INC.
XX
XX Ling V, Wu P, Gray GS;
XX
XX WPI; 2000-628257/60.
XX
XX Determining predisposition of humans to develop autoimmune disease
XX involves detecting polymorphic microsatellite repeat sequence within
XX human costimulatory receptor gene locus
XX
XX Claim 2; Page 67-82; 160pp; English.
XX
XX Two human bacterial artificial chromosome (BAC) clones that included
XX and flanked the human CTLA-4 locus were cloned and sequenced. The
XX sequence data was assembled into a contiguous sequence that is presented
XX in AAA96363-68. AAA96363-64 comprise BAC clone 22700, and AAA96365-68
XX comprise BAC clone 22608. The sequences contain polymorphic
XX microsatellite repeat (PMR) sequences. The specification describes a
XX method for determining the predisposition of a human subject to develop
XX autoimmune disease. The method comprises detecting a PMR sequence in the
XX CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
XX locus (hCGRL). PMR sequences vary in length among individuals and can be
XX amplified to generate products that differ in size. These products can
XX then be detected by rapid and convenient high resolution processes. The
XX method is useful for determining the predisposition of insulin-dependent
XX diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune

```

```

CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
CC coeliac disease and leprosy. PMR sequences within hCGRL are useful as
CC markers in a variety of assays and in the field of forensic medicine, as
CC disease diagnosis and human genome mapping.
XX
XX Sequence 50000 BP; 14521 A; 9882 C; 11191 G; 14406 T; 0 other;
SO
Query Match 4.1%; Score 20; DB 21; Length 50000;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 251 AGAGTTTGGAAGGTAAGA 270
Db 3317 AGAGTTTGGAAGGTAAGA 3336
RESULT 6
AAC48250/C
ID AAC48250 standard; DNA; 1300 BP.
XX
XX AAC48250;
AC
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 56805.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 16-APR-1999; 99US-0128714.
XX
XX 19-APR-1999; 99US-0129845.
XX
XX 21-APR-1999; 99US-0130077.
XX
XX 23-APR-1999; 99US-0130449.
XX
XX 28-APR-1999; 99US-0130510.
XX
XX 30-APR-1999; 99US-0130891.
XX
XX 30-APR-1999; 99US-0131449.
XX
XX 04-MAY-1999; 99US-0132048.
XX
XX 05-MAY-1999; 99US-0132407.
XX
XX 06-MAY-1999; 99US-0132484.
XX
XX 06-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 11-MAY-1999; 99US-0132487.
XX
XX 14-MAY-1999; 99US-0132863.
XX
XX 14-MAY-1999; 99US-0134256.
XX
XX 14-MAY-1999; 99US-0134218.
XX
XX 14-MAY-1999; 99US-0134221.
XX
XX 14-MAY-1999; 99US-0134221.
XX
XX 18-MAY-1999; 99US-0134370.
XX
XX 19-MAY-1999; 99US-0134768.
XX
XX 20-MAY-1999; 99US-0134941.
XX
XX 21-MAY-1999; 99US-0135124.
XX
XX 21-MAY-1999; 99US-0135353.
XX
XX 24-MAY-1999; 99US-0135629.
XX
XX 25-MAY-1999; 99US-0136021.
XX
XX 27-MAY-1999; 99US-0136392.

```

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-01455913.  
PR 27-JUL-1999; 99US-01455913.  
PR 27-JUL-1999; 99US-01455913.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0165596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 3.9%; Score 19; DB 21; Length 1300;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;DY 415 CAGACGACGACATTACG 433  
DB 121 CAGACGACGACATTACG 103

## RESULT 7

AAC47949/c

ID AAC47949 standard; DNA; 1302 BP.

AC AAC47949;

DN 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55704.

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 990S-0121825.

PR 09-MAR-1999; 990S-0121810.

PR 23-MAR-1999; 990S-0123548.

PR 25-MAR-1999; 990S-0125788.

PR 29-MAR-1999; 990S-0126264.

PR 01-APR-1999; 990S-0126785.

PR 08-APR-1999; 990S-0127462.

PR 16-APR-1999; 990S-0128234.

PR 19-APR-1999; 990S-0128714.

PR 21-APR-1999; 990S-0129845.

PR 23-APR-1999; 990S-0130077.

PR 28-APR-1999; 990S-0130510.

PR 30-APR-1999; 990S-0130891.

PR 04-MAY-1999; 990S-0132048.

PR 05-MAY-1999; 990S-0132407.

PR 06-MAY-1999; 990S-0132483.

PR 07-MAY-1999; 990S-0132487.

PR 11-MAY-1999; 990S-0132863.

PR 14-MAY-1999; 990S-0134256.

PR 14-MAY-1999; 990S-0134218.

PR 14-MAY-1999; 990S-0134221.

PR 18-MAY-1999; 990S-0134370.

PR 19-MAY-1999; 990S-0134768.

PR 20-MAY-1999; 990S-0134941.

PR 21-MAY-1999; 990S-0135124.

PR 24-MAY-1999; 990S-0135353.

PR 25-MAY-1999; 990S-0135629.

PR 27-MAY-1999; 990S-0136021.

PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137724.  
PR 07-JUN-1999; 990S-0138094.  
PR 08-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139110.  
PR 16-JUN-1999; 990S-0139452.  
PR 17-JUN-1999; 990S-0139453.  
PR 18-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 21-JUN-1999; 990S-0139763.  
PR 22-JUN-1999; 990S-0139817.  
PR 23-JUN-1999; 990S-0139899.  
PR 24-JUN-1999; 990S-0140353.  
PR 28-JUN-1999; 990S-0140354.  
PR 29-JUN-1999; 990S-0140695.  
PR 30-JUN-1999; 990S-0140823.  
PR 01-JUL-1999; 990S-0140991.  
PR 02-JUL-1999; 990S-0141287.  
PR 06-JUL-1999; 990S-0141842.  
PR 06-JUL-1999; 990S-0142154.  
PR 08-JUL-1999; 990S-0142055.  
PR 09-JUL-1999; 990S-0142390.  
PR 12-JUL-1999; 990S-0142920.  
PR 13-JUL-1999; 990S-0142977.  
PR 14-JUL-1999; 990S-0143542.  
PR 15-JUL-1999; 990S-0143624.  
PR 16-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 20-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 21-JUL-1999; 990S-0144632.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0145086.  
PR 22-JUL-1999; 990S-0145086.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145089.  
PR 23-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145218.  
PR 26-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145226.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 28-JUL-1999; 990S-0145918.  
PR 02-AUG-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 03-AUG-1999; 990S-0146389.  
PR 04-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 09-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.



PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151055.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159284.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 3 9%; Score 19; DB 21; Length 1302;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 415 CAGAAGCAGAGATTACG 433  
|||||

DB 123 CAGAAGCAGAGATTACG 105  
RESULT 8  
AAS26580  
ID AAS26580 standard; cDNA; 401 BP.  
XX  
AC AAS26580;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human cDNA encoding a novel secreted protein, Seq ID 759.  
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;  
KW cytoskeletal; cardiant; vasotropic; cerebroprotective; neurotropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnery; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin aging; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN M0200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225216.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227009.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.

PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250300.  
PR 01-DEC-2000; 2000US-0250360.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 06-DEC-2000; 2000US-0251988.  
PR 08-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251749.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251969.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254597.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-488783/53.  
DR P-PADB; AAI16593.  
XX

PT New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PS used as food additives or preservatives -  
PS Claim 1: SEQ ID No 759; 980pp; English.

XX  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC include autoimmunity diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence encodes a novel secreted protein of the invention.

Query Match  
Best Local Similarity 3.7%; Score 18; DB 22; Length 401;  
Matches 18; Conservative 0; Pred. No. 37;  
Mismatch 0; Indels 0; Gaps 0;

DB 384 AGAATTTCAGCAAAAGGC 401  
|||||  
380 AGAATTTCAGCAAAAGGC 397

RESULT 9  
AAFI16080  
ID AAFI16080 standard; CDNA: 932 BP.  
XX

```

AC AAF16080;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:515.
XX
KW Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytoskeletal; cardioprotective; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO20005174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PK 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM:
XX
DR WPI: 2000-587513/55.
DR P-PSDB: AAB56877.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer.
XX
PS Claim 1; Page 979; 2338bp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56863 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytoskeletal,
CC cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 932 BP; 248 A; 223 C; 211 G; 242 T; 8 other:
XX
Query Match 3.7%; Score 18; DB 21; Length 932;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 237 AATAAGCTAACCAGAG 254
DB 692 AATAAGCTAACCAGAG 709

```

```

KW aspc gene; feedback inhibition; amino acid synthesis;
XX composite plasmid; ss.
XX
OS Escherichia coli.
XX
PN WO8700202-A.
XX
PD 15-JAN-1987.
XX
PF 24-JUN-1986; 86WO-US01353.
XX
PK 24-JUN-1985; 85US-0747732.
XX
PA (NUTR-) NUTRASWEET CO.
XX
PI Edwards MR, Taylor PP, Hunter MG, Fotheringham IG;
XX
DR WPI: 1987-021998/03.
XX
DR P-PSDB: AAP70751.
XX
PT Composite plasmids contg. multiple genes in transcriptional units
XX useful for prodn. of aminoacid(s), esp. L-phenylalanine and
XX L-tyrosine
XX
PS Disclosure; Page 25; 57pp; English.
XX
CC This sequence may be inserted into a composite plasmid and used
CC for the production of amino acids. See also AAN71053-55,
CC AAN71107, AAN71109-11 and AAP70696-97 and AAP70750, AAP70752-54.
XX
SQ Sequence 1293 BP; 322 A; 331 C; 343 G; 297 T; 0 other:
XX
Query Match 3.7%; Score 18; DB 8; Length 1293;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 132 GCTGTTGCTAAGGTAG 149
DB 254 GCTGTTGCTAAGGTAG 271

```

RESULT 11  
AAV40259 standard; DNA: 1331 BP.

AC AAV40259;  
XX  
DT 13-OCT-1998 (first entry)  
XX  
DE Escherichia coli aspc gene.  
XX  
KW Brevibacterium lactofermentum; lysc; L-lysine; coryneform bacterium;  
KW aspartokinase; feedback inhibition; dihydrotoluidine reductase;  
KW diaminopimelate decarboxylase; aspartate aminotransferase; ds.  
XX  
OS Escherichia coli.  
XX  
FH Key Location/Qualifiers  
FT CDS 10..1197  
FT /tag= a  
FT /product= "aspc"  
XX  
PN EP854189-A2.  
XX  
PD 22-JUL-1998.  
XX  
PF 05-DEC-1997; 97EP-0121443.  
XX  
PK 05-DEC-1996; 96JP-0325659.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Araki M, Nakamatsu T, Sugimoto M, Yoshihara Y;

XX WPI: 1998-379060/33.  
 DR P-PSDB: AAM69553.  
 XX  
 PT Recombinant DNA autonomously replicable in coryneform bacteria -  
 PT used to produce L-lysine, codes for e.g. aspartokinase,  
 PT dihydrodipicolinate reductase and synthase and di- amino- pinelate  
 XX decarboxylase  
 PS  
 XX Example 5: Page 37-38; 59pp; English.  
 CC The present invention describes a recombinant DNA autonomously  
 CC replicable in cells of coryneform bacteria (CB) comprising a DNA  
 CC sequence coding for an aspartokinase (AK) in which feedback inhibition  
 CC by L-lysine and L-threonine is desensitized a DNA sequence coding for  
 CC dihydrodipicolinate synthase (DHPS), a DNA sequence coding for  
 CC diaminopimelate decarboxylase (DAPD), a DNA sequence coding for  
 CC aspartate aminotransferase (AAT), and a DNA sequence coding for  
 CC aspc from Escherichia coli. The present sequence encodes  
 CC by CB. The L-lysine produced can be used for improving L-lysine productivity  
 CC  
 SQ Sequence 1331 BP; 330 A; 340 C; 350 G; 311 T; 0 other;  
 Query Match 3.7%; Score 18; DB 19; Length 1331;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 132 GCGTTTGGTAAAGCTAG 149  
 DB 252 GCGTTTGGTAAAGCTAG 269

RESULT 12  
 ID AAN71109 standard; DNA; 3659 BP.  
 AC AAN71109;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE phea arof aspc operon in plasmid pME219.  
 XX  
 KW alpha-amyrase; feedback inhibition; amino acid synthesis;  
 KW composite plasmid; ss.  
 OS Bacillus licheniformis.  
 XX  
 FH Key  
 FT CDS Location/Qualifiers  
 FT 69..1241  
 FT /\*tag= a  
 FT 1259..2329  
 FT /\*tag= b  
 FT 2344..3534  
 FT /\*tag= c  
 XX  
 PN W08700202-A.  
 PD 15-JAN-1987.  
 XX  
 PF 24-JUN-1986; 86WO-US01353.  
 XX  
 PR 24-JUN-1985; 85US-0747732.  
 XX  
 PA (NUTR-) NUTRASWEET CO.  
 PI Edwards MR, Taylor PP, Hunter MG, Fotheringham IG;  
 DR WPI: 1987-021998/03.  
 DR P-PSDB: AAP70752, AAP71677-8.  
 XX  
 PT Composite plasmids contg. multiple genes in transcriptional units

PT - useful for prodn. of aminoacid(s), esp. L-phenylalanine and  
 PT L-tyrosine  
 XX  
 PS Disclosure; Page 38; 57pp; English.  
 XX  
 CC This sequence may be inserted into a composite plasmid and used  
 CC for the production of amino acids. See also AAN71053-55,  
 CC AAN71107, AAN71109, AAN71111 and AAP70696-97 and AAP70750, AAP70752 and  
 CC AAP70754.  
 XX  
 SQ Sequence 3659 BP; 936 A; 911 C; 947 G; 865 T; 0 other;  
 Query Match 3.7%; Score 18; DB 8; Length 3659;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 132 GCGTTTGGTAAAGCTAG 149  
 DB 2586 GCGTTTGGTAAAGCTAG 2603

RESULT 13  
 ID AAQ90655 standard; CDNA; 3776 BP.  
 AC AAQ90655;  
 XX  
 DT 11-NOV-1995 (first entry)  
 XX  
 DE Eph-related PTK Cdk9 cDNA.  
 XX  
 KW Cdk9; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;  
 KW prognosis; ss.  
 OS Gallus sp.  
 XX  
 FH Key  
 FT CDS Location/Qualifiers  
 FT 290..3311  
 FT /\*tag= a  
 XX  
 PN W09515375-A.  
 PD 08-JUN-1995.  
 XX  
 PF 07-SEP-1994; 94WO-US10140.  
 XX  
 PR 03-DEC-1993; 93US-0162809.  
 XX  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 XX  
 PI Pasquale EB, Sajedi FG;  
 DR WPI: 1995-215256/28.  
 DR P-PSDB: AAR75707.  
 XX  
 PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing  
 PT cancer.  
 PS  
 XX Claim 2: Page 50-53; 129pp; English.  
 CC Novel Eph-related PTK cDNA clone Cdk9 (given in AAQ90655) was  
 CC isolated from a chick embryo library in lambda. Cdk9 protein  
 CC (AAR75707) is closely related to Cdk5 (AAR75712). Cdk9 protein  
 CC Cdk9 expression is predominant in the thymus and detectable in  
 CC brain, retina, kidney, lung and heart.  
 XX  
 SQ Sequence 3776 BP; 1005 A; 881 C; 941 G; 949 T; 0 other;  
 Query Match 3.7%; Score 18; DB 16; Length 3776;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 165 GGCATTGGTGGTCTGT 182

Db 271 GGCCATTGCTGCTGT 288

## RESULT 14

AA13059 standard; DNA; 9813 BP.

AA13059;

19-MAR-1999 (first entry)

Enterococcus faecalis genome contig SEQ ID NO:122.

Enterococcus faecalis; contig; detection; Enterococcal infection;

vacine; attenuation; computer readable medium; ds.

Enterococcus faecalis.

WO985055-A2.

12-NOV-1998.

04-MAY-1998; 98WO-US08985.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

Barash SC, Dillon PJ, Kunsch CA;

WPI; 1999-045171/04.

New isolated Enterococcus faecalis polynucleotides and polypeptides

used to develop products for the detection of Enterococcus and for

infection.

Claim 1; Page 734-739; 2084pp; English.

A computer readable medium has been developed which has recorded on it

982 nucleotide sequences isolated from the Enterococcus faecalis genome.

AA13059 to AA13319 represent these nucleotide sequences which are

primary nucleotide sequences, also known as contigs. The computer-based

system can identify fragments of the Enterococcus faecalis genome with

commercial importance. The products can be used to detect the presence

of Enterococcus faecalis in samples. They can also be used for

diagnosing Enterococcal infection in an animal and monitoring

progression of disease, and for identifying agents which can be used to

modulate the growth or pathogenicity of Enterococcus faecalis, or

another related organism, in vivo or in vitro. In particular the

polypeptides encoded by the Enterococcus faecalis nucleotide sequences

can be used in vaccines to prevent or attenuate an Enterococcal

infection.

Sequence 9813 BP; 3233 A; 1698 C; 2161 G; 2719 T; 2 other;

Query Match 3.7%; Score 18; DB 20; Length 9813;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2473 ATTATATCACTGCTTT 2490

## RESULT 15

ABL10082 standard; CDNA; 10432 BP.

ABL10082;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polynucleotide SEQ ID NO 24728.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

P-PSDB; ABB65979.

New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell

interactions -

Claim 1; SEQ ID NO 24728; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is

useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of

insecticides, therapeutics and pharmaceutical drugs. The invention

discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

sequences (ABL01840-ABL16175) and the encoded proteins

(ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 10432 BP; 2894 A; 2225 C; 2170 G; 3143 T; 0 other;

Query Match 3.7%; Score 18; DB 23; Length 10432;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7018 TTGGTGTCTGCTTTTGG 7035

170 TTGGTGTCTGCTTTTGG 187

TTGGTGTCTGCTTTTGG 7035

Search completed: October 27, 2002, 20:42:22  
Job time : 142.109 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 18:35:09 ; Search time 940.383 Seconds  
(without alignments)  
10815.053 Million cell updates/sec

Title: US-09-677-374-1

Perfect score: 486  
Sequence: 1 atgaacagagatgttgcg.....aagtcattccacagaanaa 486

## Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

## Post-processing: Listing first 45 summaries

## Database :

GenEmbl:.\*  
1: gb\_da:.\*  
2: gb\_htg:.\*  
3: gb\_in:.\*  
4: gb\_cm:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vl:.\*  
15: em\_ba:.\*  
16: em\_fun:.\*  
17: em\_hum:.\*  
18: em\_in:.\*  
19: em\_mu:.\*  
20: em\_om:.\*  
21: em\_or:.\*  
22: em\_ov:.\*  
23: em\_pat:.\*  
24: em\_ph:.\*  
25: em\_pl:.\*  
26: em\_ro:.\*  
27: em\_sts:.\*  
28: em\_un:.\*  
29: em\_vl:.\*  
30: em\_htg\_hum:.\*  
31: em\_htg\_inv:.\*  
32: em\_htg\_other:.\*  
33: em\_htg\_inv:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query  
No. Score Match Length DB ID Description

1	486	100.0	489	6	AX252413	AX252413 Sequence
2	486	100.0	4983	1	AF184152	AF184152 Piscirickettsia
3	21	4.3	121922	2	AP000576	AP000576 Homo sapi
4	21	4.3	146327	9	AL162253	AL162253 Human DNA
5	21	4.3	160394	2	AC012273	AC012273 Homo sapi
6	21	4.3	161582	9	AP001636	AP001636 Homo sapi
7	21	4.3	161586	2	AC090309	AC090309 Homo sapi
8	21	4.3	165208	2	AC098770	AC098770 Homo sapi
9	21	4.3	183317	2	CNS07EFT	CNS07EFT Rattus no
10	21	4.3	191454	2	AC093459	AC093459 Homo sapi
11	21	4.3	192810	2	AL626767	AL626767 Mus muscu
12	21	4.3	252571	2	AC098728	AC098728 Mus muscu
13	21	3.875	3875	3	DP023357	DP023357 Drosophila
14	20	4.1	26058	2	AC110102	AC110102 Rattus no
15	20	4.1	28798	1	U95370	U95370 Bacillus
16	20	4.1	31147	2	AC018318	AC018318 Drosophila
17	20	4.1	32386	1	BLA05061	BLA05061 Bacillus
18	20	4.1	51409	1	AF024666	AF024666 Corynebact
19	20	4.1	72732	3	AC004715	AC004715 Drosophila
20	20	4.1	151673	10	AC087102	AC087102 Rattus no
21	20	4.1	160970	3	AC010138	AC010138 Homo sapi
22	20	4.1	180633	3	AC093549	AC093549 Drosophila
23	20	4.1	181654	9	AF411058	AF411058 Homo sapi
24	20	4.1	196247	2	AC104137	AC104137 Homo sapi
25	20	4.1	274307	3	AE003561	AE003561 Drosophila
26	19	3.9	4194	8	AJ005079	AJ005079 Kluyverom
27	19	3.9	5480	10	RATRP15	RATRP15 Rat salivar
28	19	3.9	37991	9	AP000565	AP000565 Homo sapi
29	19	3.9	39567	2	AC107960	AC107960 Homo sapi
30	19	3.9	56538	2	AC092561	AC092561 Clona sav
31	19	3.9	92693	2	AL157766	AL157766 Human DNA
32	19	3.9	100109	2	AC095401	AC095401 Rattus no
33	19	3.9	117807	2	AC106566	AC106566 Rattus no
34	19	3.9	119485	2	AC084274	AC084274 Mus muscu
35	19	3.9	121500	2	AC007916	AC007916 Homo sapi
36	19	3.9	130621	2	AC107870	AC107870 Homo sapi
37	19	3.9	131892	2	AB038490	AB038490 Homo sapi
38	19	3.9	150001	9	AC006063	AC006063 Homo sapi
39	19	3.9	160169	2	AC016660	AC016660 Homo sapi
40	19	3.9	160666	2	AC105755	AC105755 Homo sapi
41	19	3.9	167005	9	AL158070	AL158070 Human DNA
42	19	3.9	169856	9	AL591944	AL591944 Homo sapi
43	19	3.9	173646	2	AC025853	AC025853 Homo sapi
44	19	3.9	174019	2	AC025734	AC025734 Homo sapi
45	19	3.9	174130	2	AC108026	AC108026 Homo sapi

## ALIGNMENTS

RESULT 1  
AX252413  
LOCUS AX252413 489 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 5 from Patent WO016885.  
ACCESSION AX252413  
VERSION AX252413.1 GI:15985721  
KEYWORDS  
SOURCE  
ORGANISM  
Piscirickettsia salmonis.  
Piscirickettsia salmonis  
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;  
Piscirickettsia.

## REFERENCE

1 (bases 1 to 489)  
Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and  
Burzio,L.

TITLE Fish vaccine against piscirickettsia salmonis  
JOURNAL Patent: WO 016885-A 5 20-SEP-2001;  
Aqua Health (Europe) Limited (GB)

## FEATURES

source  
1..489  
Location/Qualifiers  
/organism="Piscirickettsia salmonis"  
/db\_xref="taxon:1238"

BASE COUNT 139 a 79 c 144 g 127 t  
ORIGIN

Query Match 100.0%; Score 486; DB 6; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-248;  
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGACAGAGAGATGTTGCAAGTAGTACTATATATATATCATGATGTTTACTTGGC 60
Db 1 ATGACAGAGAGATGTTGCAAGTAGTACTATATATATATCATGATGTTTACTTGGC 60
QY 61 TGTGCCAGAACTTGTAGTGTCTCAAGAGTCGACGCTGGGGCTGTGTGGCGGT 120
Db 61 TGTGCCAGAACTTGTAGTGTCTCAAGAGTCGACGCTGGGGCTGTGTGGCGGT 120
QY 121 GTTGTGGCCAGCTGTTTGTAAAGTAGTGTGTCAGTTGCAATGGCCATGTGTCCT 180
Db 121 GTTGTGGCCAGCTGTTTGTAAAGTAGTGTGTCAGTTGCAATGGCCATGTGTCCT 180
QY 181 GTTGTGGTGATTAATGTTCTTAAATCGGTCAATGATGATCAGCAGATAAATA 240
Db 181 GTTGTGGTGATTAATGTTCTTAAATCGGTCAATGATGATCAGCAGATAAATA 240
QY 241 AAGCTTAACACAGATTTGGAAAGGTAAAGCAGGGCAGAGTGGACGTTGGCGTAATCCA 300
Db 241 AAGCTTAACACAGATTTGGAAAGGTAAAGCAGGGCAGAGTGGACGTTGGCGTAATCCA 300
QY 301 GATACAGGCAATAGTTATAGTGTGAGCCAGTGGCTACTTACACGCTTACAATAAGCAA 360
Db 301 GATACAGGCAATAGTTATAGTGTGAGCCAGTGGCTACTTACACGCTTACAATAAGCAA 360
QY 361 GAGCGTCCCGCAGCAATATGTCGAAATTCAGCAAAAGGCGATGTGACGGCAGAAG 420
Db 361 GAGCGTCCCGCAGCAATATGTCGAAATTCAGCAAAAGGCGATGTGACGGCAGAAG 420
QY 421 CAAGGATTTACGGCACTGTCATGCGGCAACCGGATGTCGTGGCAATCATTTCAACA 480
Db 421 CAAGGATTTACGGCACTGTCATGCGGCAACCGGATGTCGTGGCAATCATTTCAACA 480
QY 481 GAAAAA 486
Db 481 GAAAAA 486

```

RESULT 2  
 AF184152 4983 bp DNA linear BCT 01-OCT-2000  
 LOCUS Piscirickettsia salmonis alanine racemase (alr) gene, partial cds;  
 DEFINITION and BAX (bax), DNA repair enzyme RadA (rada), 17 kDa antigen  
 (ospA), and transposase (tnpA) genes, complete cds.  
 AF184152  
 AF184152.1 GI:10441343

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 SOURCE  
 Piscirickettsia salmonis.  
 Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;  
 Piscirickettsia.  
 1 (bases 1 to 4983)  
 Kuzuk, M.A., Burlan, J., Thornton, J.C. and Kay, W.W.  
 Identification of a genus-common Rickettsial surface antigen in the  
 salmonid pathogen Piscirickettsia salmonis  
 Unpublished  
 2 (bases 1 to 4983)  
 Kuzuk, M.A. and Kay, W.W.  
 Direct Submission  
 Submitted (10-SEP-1999) Biochemistry & Microbiology, University of  
 Victoria, Peach Building, PO Box 3055, Victoria, BC V8W 3P6, Canada  
 Location/Qualifiers  
 1. 4983  
 /organism="Piscirickettsia salmonis"  
 /strain="JF-89"  
 /db\_xref="taxon:1238"  
 /clone="PB12"  
 <1..501  
 /gene="alr"

gene

CDS

gene

CDS

gene

CDS

source

```

<1..501
/gene="alr"
/codon_start=1
/transl_table=11
/evidence-not-experimental
/product="alanine racemase"
/protein_id="AAG17001.1"
/db_xref="GI:10441347"
/transl_table="NSAIFNFSERYDWRPGIMLYGISPADKNCVDLELOPVNHV
VSRLISVROLQSGESVYGATWOCPEDMOYGLISGDCYPRLASGTFPLVGGRC
PYMOE"
complement(526..1257)
/gene="bax"
complement(526..1257)
/gene="bax"
/codon_start=1
/transl_table=11
/evidence-not-experimental
/product="BAX"
/protein_id="AAG17002.1"
/db_xref="GI:10441348"
/transl_table="MSLAVITTAQAQPLFSAPTSKTOINTAVSODIOSIKOFEIT
KITQSENTIMOILQKROHILSLQKSSKEITQOLALQAYOLNNHREHPDPS
LLKRVDIIPNALISAQAINESNMGSRSAVGNPFGRCHTPEGGCIIPARPNH
WEVANYSSMTASVRAVIHTLINTHNAVQALRDLRAHMRANHDVSAFKLAEGLTAYSIK
1373..2743
/gene="rada"
/evidence-experimental
1373..2743
/gene="rada"
/codon_start=1
/transl_table=11
/product="DNA repair enzyme RadA"
/protein_id="AAG16998.1"
/db_xref="GI:10441344"
/transl_table="WSKQNOYVCSDCGIAKKMLGOCPHCOQNSLSSEVKYLPNP
GNSQRPAGFAGIDAPKVCLEITPEQISRPQVGCEDRVLAGGIVHGAIVILGDP
GKSTLLLOTNSVCTOPKVLVYTGESLEQVTLRSKRIGLSQDVNLRLLAEITQVR
ILKAAEIQPKVLIVDSIQITFTESLOSAPGVSAOVRESAAILIQAKRTGCLPLVG
HYTKEGALAGPVLHEHWMDVLYFEGBDSPFLRLRAVKNFEGANSLGIFAMTEGL
KTVSNPAIFLSRYEDLOPGSVYAMGTRPFLVEVQALVDESHSNPRLIAGVQD
ORLAMLAVINRHGSIASYDDVFTINVGKKTITETADALAILLACVSLRGKALSOE
LIVEPEIGLSEIRPVQGRLEIRKEAVKHGFTQAIPLANCPKQKVGIRIVGKMLE
DALATL..3322
/gene="ospA"
/evidence-experimental
2834..3322
/gene="ospA"
/note="similar to the rickettsial 17 kDa antigen"
/codon_start=1
/transl_table=11
/product="17 kDa antigen"
/protein_id="AAG17000.1"
/db_xref="GI:10441346"
/transl_table="MNRGCLQSSLLIISVFLVGCQAQNFROEVAATGAVGVYAGQ
LFGVSGRVAMATGAVLGLISKIGQSMDOQDKIKMOSLEKVKAGQVTRMRPDT
EK"
3560..4849
/gene="Piscirickettsia salmonis"
/strain="JF-89"
/db_xref="taxon:1238"
/clone="PB12"
/insertion_seq="insertion sequence-like element"
3560..3846
/gene="tnpA"
complement(3842..4561)
/gene="tnpA"
/codon_start=1

```



```
repeat_region
BASE COUNT      1362 a      778 c      1265 g      1578 t
ORIGIN
      /transl_table=1
      /product="transposase"
      /protein_id="FAG16999.1"
      /db_xref="GI:1041345"
      /translation="MDLTLLSLFCVIDDFCOELLPPONNAILEDYTNKRNRKPKSPOMSTTS
      EIMTMTYLFPHKSNYRNKPKMYLLHVAKGSMVKYFNPNSVYNRFELMSLILCFEFA
      AOGRTATGIVFVDSITILRVCHERKASNSAPFKGLAKSKSTMGWYGFGLKIIIVNDGGA
      ALMEFKMSKATTDYRVLLPKMAENLTUKIGDKVYISQKLFDOLEYGLOLITIRRNK
      MNKLVLMDIKRILLKRRGIIIA"
      4562      4849
```

```
repeat_region 4562. 4849
BASE COUNT    1362 a    778 c    1265 g    1578 t
ORIGIN
```

Query Match	100.0%; Score 486; DB 1; Length 4983;
Best local similarity:	100.0% ; Prod vs A 66-748.

QY	1	ATGAAACAGAGAGATGGTTTGGCAAGGTACTACTCTAAATATATATACAGTGTGTTTTAACTGGC	60
Db	2834	ATGAAACAGAGAGATGTTTGCAGAGGTACTACTCTAAATATATATATACAGTGTGTTTTAACTGGC	2893
QY	61	TGTGCCAGAACTTTTACTGCTCAAGAATCGGAGCTGGCAGCTGGGGCTGTGTGGCGGT	120
Db	2894	TGTGCCAGAACTTTTACTGCTCAAGAATCGGAGCTGGCAGCTGGGGCTGTGTGGCGGT	2953
QY	121	GTTGCTGCGCAGCTGTTTGGTAAAGTACTGGTCGAGCTTGCAATGAGCCATTGGTGGTGTCT	180
Db	2254	GTTGCTGCGCAGCTGTTTGGTAAAGTACTGGTCGAGCTTGCAATGAGCCATTGGTGGTGTCT	3013
QY	181	GTTTTGGGTGATTTAATTGGTTCTAAATCGGTCCAATCGATGGATCAGCAGGATAAATA	240
Db	3014	GTTTTGGGTGATTTAATTGGTTCTAAATCGGTCCAATCGATGGATCAGCAGGATAAATA	3073
QY	241	AAGCTAAACAGAGTTTGGAAAAGGTAAAGACAGGGCAAGTACACGTTGGCGTAAATCA	300
Db	3074	AAGCTAAACAGAGTTTGGAAAAGGTAAAGACAGGGCAAGTACACGTTGGCGTAAATCA	3133
QY	301	GATACAGCAATAGTTATAGTGTGAGCCAGTGGCTACTTACACAGCTTACATTAAGCA	360
Db	3134	GATACAGCAATAGTTATAGTGTGAGCCAGTGGCTACTTACACAGCTTACATTAAGCA	3193
QY	361	GAGCGTGGCAGCAATATTGTGCAAGATTTCAGCAAAAGCCATGATTGCAAGGCAGAG	420
Db	3194	GAGCGTGGCAGCAATATTGTGCAAGATTTCAGCAAAAGCCATGATTGCAAGGCAGAG	3253
QY	421	CAAGAGATTTACGGCACTGCATCCCGCAACCGGATGTGCTTTGGCAAGTCATTTCAACA	480
Db	3254	CAAGAGATTTACGGCACTGCATCCCGCAACCGGATGTGCTTTGGCAAGTCATTTCAACA	3313
QY	481	GAATAA 486	
Db	3314	GAATAA 3319	

[illegible]

TITLE  
JOURNAL

**COMMENT**

Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y. Direct Submission  
Submitted (08-OCT-1999) Masahira Hattori, The Institute of Physical and  
Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-15-1, Katsushika, Sagamihara, Kanagawa 228-8555, Japan  
(E-mail: hattori@gsc.riken.go.jp, URL: <http://gsc.riken.go.jp/>;  
Tel: 81-42-778-9923, Fax: 81-42-778-9924)  
On Aug 26, 2000 this sequence version replaced g1:8118784.

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      25319 contig of 25319 bp in length
25420  45356 contig of 18117 bp in length
43637  62163 contig of 18527 bp in length
62264  74470 contig of 12207 bp in length
74571  84826 contig of 10256 bp in length
84927  94411 contig of 9455 bp in length
94512  103560 contig of 9049 bp in length
10361  111843 contig of 8183 bp in length
111944  116839 contig of 4896 bp in length
116840  116930 gap of 100 bp
116940  116947 contig of 2708 bp in length
119648  119747 gap of 100 bp

NOTE: This is a 'working draft' sequence. It currently
consists of 12 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1      25319 contig of 25319 bp in length
25320  25419 gap of 100 bp
25420  43536 contig of 18117 bp in length
43637  43636 gap of 100 bp
43637  62163 contig of 18527 bp in length
62164  62263 gap of 100 bp
62264  74470 contig of 12207 bp in length
74471  74570 gap of 100 bp
74571  84826 contig of 10256 bp in length
84827  84926 gap of 100 bp
84927  94411 contig of 9485 bp in length
94412  94511 gap of 100 bp
94512  103560 contig of 9049 bp in length
10361  103660 gap of 100 bp
10361  111843 contig of 8183 bp in length
111844  111943 gap of 100 bp
111944  116839 contig of 4896 bp in length
116840  116930 gap of 100 bp
116940  116947 contig of 2708 bp in length
119648  119747 gap of 100 bp

```

	Query Match	4.3%	Score 21:	DB 2;	Length 121922;	
	Best Local Similarity	100.0%;	Pred. No. 5;			
	Matches 21:	Conservative 0;	Mismatches			
OY	34 ATTATTATTCAGTGGCTTTTA	54		0;	Indels 0;	Gaps 0;
Db	70697 ATTATTATTCAGTGGCTTTTA	70717				

REFERENCE  
1 (bases 1 to 146327)  
AUTHORS  
TITLE  
Catarrhini, Homidae, Homo.  
Direct Submission  
Submitted (70-nu)

requests: 30nerequests@anger.ac.uk  
 requests: 30nerequests@anger.ac.uk  
 On Apr 19, 2001 this sequence version replaced 91:12539553.  
 During difference assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone. Note that the  
 only a small overlap as described above, as we submit sequences with  
 This sequence was finished as follows: unsequenced regions were

```
misc_feature misc_feature
```

RESULT	5	AC012273	LOCUS	DEFINITION	AC012273	Accession	Version	Keywords	Source	Organism
					AC012273	Human sapiens clone RP11-1K7, WORKING DRAFT SEQUENCE, 4 unordered pieces.	AC012273.3	GI:14595892	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.
					AC012273	Human sapiens clone RP11-1K7, WORKING DRAFT SEQUENCE, 4 unordered pieces.	AC012273.3	GI:14595892	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	human.

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS	1 (bases 1 to 160394) Birren, B., Linton, L., Nusbaum, C. and Lander, E. Homo sapiens, clone RP11-1K7 Unpublished 2 (bases 1 to 160394) Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Brown, A., Castle, A., Colangelo, R., Boguski, L., Bouckgeater, B., Cocke, P., DeBellano, R., Dewar, K., Collins, S., Collins, A., Fierella, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
--	--

```

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehocky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,D.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced gi:13940700.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L1474
Center Clone name: L_K_7
----- Summary Statistics
Sequencing vector: pLJM3: M77815; 24% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159424 bases at least Q40
Consensus quality: 159851 bases at least Q30
Consensus quality: 160001 bases at least Q20
Insert size: 162000; agarose-ftp
Insert size: 160094; sum-of-contigs
Quality coverage: 14.6 in Q20 bases; agarose-ftp
NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 16762: contig of 16762 bp in length
*
* 16763 16862: gap of 100 bp
*
* 16863 20224: contig of 3362 bp in length
*
* 20225 20324: gap of 100 bp
*
* 20325 97675: contig of 77351 bp in length
*
* 97676 97775: gap of 100 bp
*
* 97776 160394: contig of 62619 bp in length.
Location/Qualifiers
1. 160394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-1K7"
/clone_1fb="RP11-11 Human Male BAC"
1. 16762
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
16863..20224
/note="assembly_fragment"
20325..97675
/note="assembly_fragment"
97776..160394
/note="assembly_fragment"
clone_end:T7
vector_side:right"

BASE COUNT 49405 a 32013 c 31371 g 47305 t 300 others
ORIGIN
Query Match 4.3%; Score 21; DB 2; Length 160394;

Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 AAAGCTAAACCCAGACTTTGGA 260
|||||
Db 81250 AAAGCTAAACCCAGACTTTGGA 81270

RESULT 6
AP001636/c
LOCUS 161582 bp DNA linear PRI 08-DEC-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-142C4,
complete sequence.
ACCESSION AP001636
VERSION AP001636.4 GI:17425232
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-142C4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 161582)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gs.riken.go.jp, URL:http://nyp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Dec 7, 2001 this sequence version replaced gi:9927280.
FEATURES
SOURCE
1. 161582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-142C4"

BASE COUNT 48492 a 31636 c 32205 g 49249 t
ORIGIN
Query Match 4.3%; Score 21; DB 9; Length 161582;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 ATTATATCACTGCTTTTGA 54
|||||
Db 81922 ATTATATCACTGCTTTTGA 81902

RESULT 7
AC090309/c
LOCUS 161586 bp DNA linear HTG 01-AUG-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-142C4 map 11, *** SEQUENCING
IN PROGRESS ***; 1 ordered pieces.
ACCESSION AC090309
VERSION AC090309.4 GI:15055316
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULITOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 161586)
AUTHORS Birren,B., Linton,L., Nussbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-142C4
JOURNAL Unpublished
2 (bases 1 to 161586)
REFERENCE
Bairn,N., Bastien,V., Boguslavsky,L., Bouhgalier,B., Brown,A.,
Camarata,J., Campopiano,A., Chospel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,D.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

```

Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,  
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,  
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKenna, K.,  
 Mphahlele, R., Mphahlele, J., Mphahlele, M., Mphahlele, T.,  
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phukhng, P., Pierre, N., Pollara, V., Raymond, C., Retter, M.,  
 Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,  
 Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,  
 Sognez, C., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
 Tlivers, M., Travis, N., Triggillo, J., Vassiliev, H., Videl, R.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 1, 2001 this sequence version replaced g1:14595864.  
 All repeats were identified using RepeatMasker:  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MTBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: 142\_C\_4  
 Center Clone name: 142\_C\_4

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved  
 \* 1 161586 contig of 161586 bp in length.  
 \* Location/Qualifiers  
 1. 161586  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="11"  
 /clone="RP11-142C4"  
 /clone\_id="RP11-142C4"  
 /map="11"  
 /chromosome="11"

BASE COUNT 48496 a 31637 c 32204 g 49249 t  
 ORIGIN

Query Match 4.3% Score 21; DB 2; Length 161586;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 34 ATTATATACAGTCGTGTTT 54  
 Db 81924 ATTATATACAGTCGTGTTT 81904

RESULT 8  
 AC098770/c 165208 bp DNA linear HTG 20-DEC-2001  
 LOCUS Rattus norvegicus clone CH230-113M1, \*\*\* SEQUENCING IN PROGRESS  
 AC098770  
 AC098770.2 GI:117974352  
 HTG: HTGS PHASE1.  
 KEYWORDS Norway rat  
 SOURCE Rattus norvegicus  
 ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 165208)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., All-ouman, F.R., Allen, C.,  
 Alstbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Babaria, J.,  
 Benton, J., Blum, K., Blankenbush, K., Bonin, D., Bouck, J.,  
 Bowe, S., Britva, M., Brown, E., Brown, M., Bryant, D., Bunk, J.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.P.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G.,  
 Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dabrone, S.R., David, R., Davila, M.L., Davis, C.,  
 Day, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Dunn, A.L., Ding, Y., Din, H.H., Douthett, K.J., Drepper, H.,  
 Elhaj, C., Escotto, M., Falls, T., Farraguto, D., Edwards, C.C.,  
 Foster, P., Frenet, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolyet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, D., King, L., Kovach, J.,  
 Kovar, C., Kratovic, J., Kresh, S., Landry, N., Leal, B., Lewis, L.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,  
 Louisedge, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S.,  
 Ough, M., Okunolu, G., Oragunye, N., Oyedero, R., Pace, A., Payton, B.,  
 Peery, J., Petrucci, L., Peters, L., Pickens, R., Prins, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rivers, M., Rojas, A., Robubokun, I., Rolfe, M.,  
 Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitani, N.,  
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Swatek, A., Taylor, P., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Thomas, A., Thomas, K.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalona, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Williamson, A.,  
 Wallington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S.,  
 Wolley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 165208)  
 Morley, K.C.  
 Direct Submission  
 Submitted (01-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced g1:16572814.

## COMMENT

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc@bcm.tmc.edu  
 Project Information  
 Center project name: G17J  
 Center clone name: CH230-113M1  
 Summary Statistics  
 findphrapblast  
 Consensus quality: 136180 bases at least Q40  
 Consensus quality: 144699 bases at least Q30  
 Consensus quality: 151851 bases at least Q20  
 Estimated insert size: 138535; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agorose-1p estimation  
 Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 62 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as

[illegible]

Query Match	Best Local Similarity	4.3%: Score 21: DB 2: Length 165208;
Matches	21: Conservative	0: Mismatches 0: Indels 0: Gaps 0:
Qy 235 AAAATAAGCTAACACGAGT 255		
Db 37871 AAAATAAGCTAACACGAGT 37851		
RESULT 9		
CNS07EFI/C		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
CNS07EFI	183317 bp	DNA linear HTG 05-FEB-2001
Oryza sativa chromosome 12 clone OSUNBA0009F13, *** SEQUENCING IN		
PROGRESS ***		
AL513003		
AL513003.1 GI:12329144		
HTG: HTGS_PHASE2.		
Oryza sativa		
Oryza sativa		
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
Ehrhartoideae; Oryzaceae; Oryza.		
1 (bases 1 to 183317)		
Salse,J., Cholsie,N., Orjeda,G., Regad,F., Lorieux,M., Cooke,R.,		
Delseray,M., Robert,C., Brottier,P., Wincker,P., Cruaud,C.,		
Artiguenave,F., Saurin,W., Salanoubat,M., Quetier,F. and		
Weissenbach,J.		
Oryza sativa chromosome 12 sequencing		
Unpublished		

REFERENCE 2 (bases 1 to 183317)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-FEB-2001) Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

## COMMENT

- Web : www.genoscope.cns.fr  
 IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence  
 Work on the sequence is in progress and the release of this data is from E.coli, yeast, vector, phage, etc.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

## FEATURES

Location/Qualifiers

1..183317  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:4530"  
 /chromosome="12"  
 /clone="OSJNBa009P13"  
 /clone\_lib="CGI Nipponbare BAC"

## BASE COUNT

52033 a 38000 c 38408 g 54873 t 3 others

## Query Match

Best Local Similarity 100.0%; Score 21; DB 2; Length 183317;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 TTGGAAAAGCTAAACGACGG 276  
 Db 67144 TTGGAAAAGCTAAACGACGG 67124

## RESULT 10

AC093459/c

## LOCUS

AC093459 191454 bp DNA linear HTG 07-FEB-2002

## DEFINITION

Homo sapiens chromosome 2 clone RP11-499M11, WORKING DRAFT

## ACCESSION

AC093459

## VERSION

AC093459.4 GI:18584937

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

## AUTHORS

Waterston, R.H.

## TITLE

The sequence of Homo sapiens clone

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 191454)

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submission

## JOURNAL

Submitted (25-AUG-2001) Genome Sequencing Center, Washington

## COMMENT

MO 63108, USA  
 On Feb 7, 2002 this sequence version replaced g1.18308909.

Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.wustl.edu  
 Project Information  
 Center project name: H.NH0499M11  
 Summary Statistics  
 Sequencing vector: M13, 58  
 Sequencing method: plasmid, 948  
 Chemistry: Dye primer ET, 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: PHRAP; version 0.990319  
 Consensus quality: 180361 bases at least Q40  
 Consensus quality: 180903 bases at least Q40  
 Consensus quality: 191116 bases at least Q30  
 Insert size: 184000; agarose-fp  
 Insert size: 191254; sum-of-contigs  
 Quality coverage: 8.23 in Q20 bases; agarose-fp  
 Quality coverage: 7.92 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

## FEATURES

1..191454  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone="RP11-499M11"  
 1..9926  
 /note="assembly\_name:Contig1"  
 10027..75543  
 /note="assembly\_name:Contig17"  
 75644..191454  
 /note="assembly\_name:Contig18"

misc\_feature 1..9926

misc\_feature 10027..75543

misc\_feature 75644..191454

BASE COUNT 57526 a 36639 c 38791 g 58298 t 200 others

## Query Match

Best Local Similarity 100.0%; Score 21; DB 2; Length 191454;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 AAAGCTAAACGACGTTTGA 260  
 Db 127210 AAAGCTAAACGACGTTTGA 127190

## RESULT 11

AL626767/c

## LOCUS

AL626767 192810 bp DNA linear HTG 12-OCT-2001

## DEFINITION

Mus musculus chromosome 4 clone RP23-421C7, \*\*\* SEQUENCING IN

## ACCESSION

AL626767

## VERSION

AL626767.2 GI:16073773

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## TITLE

Direct Submission

## JOURNAL

Submitted (10-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,

## COMMENT

On Oct 11, 2001 this sequence version replaced g1.16031460.

Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 Project Information  
 Center project name: BM421C7

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 187761 bases at least Q40
Consensus quality: 188634 bases at least Q40
Consensus quality: 189284 bases at least Q20
Insert size: 190910; sum-of-contigs
Insert size: 213770; 3.4% error; agarose-fp
Quality coverage: 8.22x in Q20 bases; sum-of-contigs quality
Coverage: 7.45x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source      Location/Qualifiers
1..192810
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-421C7"
/clone_11b="RPC1-23"
1..15994
/note="assembly_fragment:02908
fragment_chain:1
clone_end:SP6
vector_side:left"
16095..26569
/note="assembly_fragment:00217
fragment_chain:1"
26670..52345
/note="assembly_fragment:00503
fragment_chain:1"
52446..57070
/note="assembly_fragment:03381
fragment_chain:1"
57171..60077
/note="assembly_fragment:02852
fragment_chain:1"
60178..62198
/note="assembly_fragment:02444
fragment_chain:1"
62289..66215
/note="assembly_fragment:00203
fragment_chain:2"
66316..93884
/note="assembly_fragment:03792
fragment_chain:2"
93985..103843
/note="assembly_fragment:03646
fragment_chain:2"
103944..116303
/note="assembly_fragment:05128
fragment_chain:2"
116404..120297
/note="assembly_fragment:00677
fragment_chain:3"
120398..124934
/note="assembly_fragment:02132
fragment_chain:3"
125035..128604
/note="assembly_fragment:03420
fragment_chain:3"
128705..131472
/note="assembly_fragment:03428
fragment_chain:4"
131573..145198
/note="assembly_fragment:04437
fragment_chain:4"
145299..148627
/note="assembly_fragment:05331"
148728..150958

```

---

```

Query Match      4.3%; Score 21; DB 2; Length 192810;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 171 TGGTGCTCTGTTTGGCTGC 191
Db 130405 TGGTGTCTCTGTTTGGCTGC 130385

RESULT 12
AC098728/c 252571 bp DNA linear HTG 14-NOV-2001
LOCUS Mus musculus chromosome UNK clone RP23-3H22, WORKING DRAFT
DEFINITION
ACCESSION AC098728
VERSION AC098728.2 GI:169241170
KEYWORDS HTG; HTGS_Phasel; HTGS_DRAFT; HTGS_ActivePtn.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 252571)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 252571)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Nov 14, 2001 this sequence version replaced gi:16554425.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BA0003H22
----- Summary Statistics -----
Sequencing vector: M13; %
Sequencing vector: plasmid: %
Chemistry: Dye-terminator Big Dye; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 249548 bases at least Q40
Consensus quality: 250157 bases at least Q30
Consensus quality: 250759 bases at least Q20
Insert size: 250000; agarose-fp
Insert size: 251971; sum-of-contigs
Quality coverage: 17.62 in Q20 bases; agarose-fp
Quality coverage: 17.54 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1055: contig of 1055 bp in length  
 \* 1056 1155: gap of unknown length  
 \* 1156 2391: contig of 1236 bp in length  
 \* 2392 2491: gap of unknown length  
 \* 2492 3645: contig of 1154 bp in length  
 \* 3646 3745: gap of unknown length  
 \* 3746 4905: contig of 1160 bp in length  
 \* 4906 5005: gap of unknown length  
 \* 5006 23951: contig of 18946 bp in length  
 \* 23952 130884: contig of 106833 bp in length  
 \* 130885 130984: gap of unknown length  
 \* 130985 252571: contig of 121587 bp in length.

FEATURES  
 source  
 1.252571  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="UNK"  
 /clone="RP23-3H22"  
 1. 1055  
 /note="assembly\_name:Contig27"  
 misc\_feature  
 1156..2391  
 /note="assembly\_name:Contig28"  
 misc\_feature  
 2492..3645  
 /note="assembly\_name:Contig35"  
 misc\_feature  
 3746..4905  
 /note="assembly\_name:Contig37"  
 misc\_feature  
 5006..23951  
 /note="assembly\_name:Contig38"  
 misc\_feature  
 24052..130884  
 /note="assembly\_name:Contig39"  
 misc\_feature  
 130985..252571  
 /note="assembly\_name:Contig40"

BASE COUNT 73702 a 54784 c 54357 g 69126 t 602 others  
 ORIGIN

Query Match 4.3% Score 21; DB 2; Length 252571;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 25 AGTAGTCTAATTTATTCAGT 45  
 Db 140043 AGTAGTCTAATTTATTCAGT 140023

RESULT 13  
 DPU22357/c 3875 bp DNA linear INV 17-APR-1997  
 LOCUS Drosophila pseudobscura runt gene, complete cds.  
 DEFINITION 022357  
 ACCESSION 022357.1 GI:722344  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Drosophila pseudobscura.  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 3875)  
 Pepling,M.E. and Gerger,J.P.  
 Conservation and function of the transcriptional regulatory protein  
 Runt  
 Proc. Natl. Acad. Sci. U.S.A. 92 (20), 9087-9091 (1995)  
 96016114  
 2 (bases 1 to 3875)  
 Pepling,M.E.  
 Direct Submission  
 Submitted (07-MAR-1995) Melissa E. Pepling, Biochemistry and Cell  
 Biology, SUNY at Stony Brook, 346 Life Sciences Building, Stony

Brook, NY 11794, USA  
 location/Qualifiers  
 1..3875  
 /organism="Drosophila pseudobscura"  
 /db\_xref="taxon:7237"  
 gene  
 prim\_transcript 1368..3875  
 join(1799..2549,2841..3775)  
 /gene="runt"  
 CDS  
 join(1799..2549,2841..3775)  
 /gene="runt"  
 /codon\_start=1  
 /protein\_id="AAB53378.1"  
 /db\_xref="GI:722345"  
 /translation="MHLPNGPTWYANSTAVAHHTOVLAAAAAANSTPAPAFSA  
 AOSTSLANTSTHSASSSTGSDTPEISATNTMTNNTSSSSNSNTNNTANNSSNSN  
 SNSNSAKPSSMTDMFASLHEMLQEVGELAQGSILCSALPHWRSNKSLPGAFK  
 VIALDDVDGTLVIRKCGNDENCGELRNCTITMKQVAKFNDLRFVRSRGSKSEFL  
 TITATYVPOIASYAKIKVTYVGPPEPRKOSGYDPHGAFNPMLNPAMDAAVMT  
 YGVADYPRHOAAQAAVHHVPAISKASPSSSSVSPASAAAGNGAGAPADYTH  
 VSGTTPPSSAPTRAGPAAWPPSPGAAVPAVATPQPFNVHAAAAAAGGACQOH  
 AKSTPAHFHPYNFAAAGLRARNAAAVLAHGGADATGISIPASRPSSTPQOH  
 LKLNUTSIETSSIHQASDADSDDDQIDQIVASEVDLXKSIDSSSFLAQHSVPLR  
 MRCDLKAPSAIKPLFHESSAAATNRQASPEITLLPAATKLNSTVQOKTWBPY"

BASE COUNT 895 a 1207 c 1002 g 766 t 5 others  
 ORIGIN

Query Match 4.1% Score 20; DB 3; Length 3875;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 108 TGTTGTTGGCGGTGTCCTG 127  
 Db 2075 TGTTGTTGGCGGTGTCCTG 2056

RESULT 14  
 AC110102 26058 bp DNA linear HTG 10-FEB-2002  
 LOCUS AC110102  
 DEFINITION Rattus norvegicus clone CH230-209C7, \*\*\* SEQUENCING IN PROGRESS  
 AC110102  
 ACCESSION \*\*\* 18 unordered pieces.  
 VERSION AC110102.1 GI:18643397  
 KEYWORDS HTG; HTGS; PHASEL.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 26058)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Allisman,F.R., Allen,C.,  
 Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbara,J., Benton,J., Bimaye,K., Blankenburg,K., Bonini,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinb,H.H.,  
 Donthwaite,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Farnhart,C., Flegg,N., Ford,J., Foster,P., Franz,P.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,  
 Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Hayida,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,  
 Homsi,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C.,  
 Kratovic,U., Kureshi,A., Landry,N., Leal,B., Lewis,J.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,



Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M.  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokinkwo, S., Ogih, M., Okunnu, G.,  
 Oregunye, N., Overledo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Plinius, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshatari, N., Sisson, I.,  
 Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Taboi, P., Tameritsa, A., Tameritsa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telirod, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gdbbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 26058)  
 Worley, K. C.  
 Direct Submission  
 Submitted (10-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center

```

Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

-----
Project Information
Center project name: GOGC
Center clone name: CH230-Q09C7
Summary Statistics
Sequencing vector: Plasmid; M77189
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 15687 bases at least Q40
Consensus quality: 17509 bases at least Q30
Consensus quality: 19045 bases at least Q20
Estimated insert size: 16607; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation
-----

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1      1634: contig of 1634 bp in length
1735   1734: gap of unknown length
1735   3200: contig of 1466 bp in length
3201   3300: gap of unknown length
3301   4502: contig of 1202 bp in length
4503   4602: gap of unknown length
4603   6078: contig of 1476 bp in length
6079   6178: gap of unknown length
6179   7942: contig of 1764 bp in length
7943   8042: gap of unknown length
8043   9277: contig of 1235 bp in length
9278   9377: gap of unknown length
9378   10585: contig of 1208 bp in length
10586  10685: gap of unknown length
10686  11995: contig of 1310 bp in length
11996  12095: gap of unknown length
12096  13724: contig of 1629 bp in length
13725  13824: gap of unknown length
13825  15152: contig of 1328 bp in length
15153  15152: gap of unknown length

```

FEATURES	source
*	15253 16751: contig of 1499 bp in length
*	16752 16851: gap of unknown length
*	16852 18107: contig of 1256 bp in length
*	18108 18207: gap of unknown length
*	18208 19468: contig of 1261 bp in length
*	19469 19568: gap of unknown length
*	19569 21159: contig of 1591 bp in length
*	21160 21259: gap of unknown length
*	21260 22506: contig of 1247 bp in length
*	22507 22606: gap of unknown length
*	22607 23692: contig of 1086 bp in length
*	23693 23792: gap of unknown length
*	23793 24867: contig of 1075 bp in length
*	24868 24967: gap of unknown length
*	24968 26058: contig of 1091 bp in length.
	Location/Qualifiers
	1. .26058
	/organism="Rattus norvegicus"
	/db_xref="taxon:10116"
	/clone="CH230-209C7"
BASE COUNT	6908 a 5623 c 5111 g 6656 t 1760 others
ORIGIN	
Query Match	4.1%; Score 20; DB 2; Length 26058;
Best Local Similarity	100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
Oy 229 CAGGATAAATTAAGCTAAA 248	
Db 19180 CAGGATAAATTAAGCTAAA 19199	

RESULT	15
LOCUS	U95370                28798 bp    DNA           linear      BCI_14-JAN-1996
DEFINITION	Bacillus licheniformis lichenysin biosynthesis operon; lichenysin synthetase A ( <i>licA</i> ), lichenysin synthetase B ( <i>licB</i> ), lichenysin synthetase C ( <i>licC</i> ), and thioesterase ( <i>liceT</i> ) genes, complete cds.
ACCESSION	U95370
VERSION	U95370.1
KEYWORDS	GI:3080741
SOURCE	.
ORGANISM	Bacillus licheniformis. Bacillus licheniformis Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus. 1 (bases 1 to 28798) Konz.D., Doeckel,S. and Marehiel,M.A. Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide Lichenysin J. Bacteriol. 181 (1), 133-140 (1999)
REFERENCE	2 (bases 1 to 28798) Konz.D., Doeckel,S. and Marehiel,M.A. Direct Submission Submitted (27-MAR-1997) Fachbereich Chemie/Biochemie, Philipps-universitaet Marburg, Hans-Meerwein-Str., Marburg 35032, Germany
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
FEATURES	
source	Location/Qualifiers 1..28798 <code>/organism=Bacillus licheniformis"</code> <code>/strain=ATCC_10716"</code> <code>/db_xref=ATCC:10716"</code> <code>/db_xref=taxon:1402"</code> <code>/note=Lichenysin biosynthesis operon"</code> 462..11210 <code>/gene=llica"</code> 462..11210 <code>/gene=llica"</code> <code>/function="antibiotic-biosynthesis"</code> <code>/note=pica"</code> <code>/codon_start=1</code> <code>/transl_table=11</code> <code>/product="lichenysin synthetase A"</code>
CDS	
gene	

```
/protein_id="AAD04757.1"
/db_xref="GI:3080742"
/translation="MGNTFFPLTHAORLIWYERKEYPGTSVSNLSGFKKSKSGTDS
GLTEAIRQFVRIETNREPRFLMEFGEDEKROYIADEPFOIECPDASRSGAGVGLKW
GOAEARRLPLPGSPLEKFAVYVNISEESMEFKAVKHIIISGISMTHILGRITDIYK
IOLDAFSEKISIPALYKTIQYCKEENHISVLSLELSILHITMHRITGQKDVLTGM
DLRERPHVKSLPSVLEYOVMOQRKETSLELPERFSSGVENDISIVKRMPTG
LAIDEPREBELTCEBELAVIERLMTLLLEALLLPKOTIALELIVPAEKELRRGTS
SOTIAYDKMNTLAGLEOKADHEKTAAYVEGOKISTRELNEOSRLARLRGRTG
PDAPAVVNERSEKIVITAMLCVAKAGAYVPIDPGEERIFRMEDEKVAIVITDGS
LVRETAEYVRSSEALSSEENGHSPSGAGHAIYIYSGTGRPKVMIEHQVH
LYNGLOOAVGADDDLLALLAFPHRDASVOQITFSLIYSGTGRPKVMIEHQVH
SOYVRRHODIVTDGPAHLOLLASADLSGVKIRMLVGEELISVATERLLDLEFAT
AESYDPVINYGPTEVCVDASSFTMNHNDLGQDRAVPIGRPIGNRPITILLENAL
LDPVEGELYAGDVGGRYLINPDITADKLEDFEPVGGMYRTGAYMWLDGTVD
FTRKDQVRIKRGRIELGEBESVLOQAPAVAGAVLARPEPSSLEVCAYVVKOSG
EHLQGLREHLSKHLDPYNTPSHFVELEIPLTSGKVDKRALRHEVSAGVETAA
FRNEYEEKVGIWQEVLTGBOVGHIDOFPELGSIKAMMLAKIKAGVEVPLQVLA
PEKPTVAALSGVSEAKDGFYIEPASEDDYPLISAORITIVSOLDAVGVNMP
AAALTEGLDVORLEAKFOKLIIDREHAIKRTSETVDEEPRONITHOVOKIEKVA
KPIEOMAKSFVRRTDLAKAPIRAGIVLADRRHVLFPDHHLVSDGVSITILNEAL
ALYGERLEPNIHTKDYAFWORTHAPGOKREAVIWOVSFAELPVQILTIDPERG
VOSEGEDRYSAVLPODMKEKTLAVLQNGATLYMWALVNMULAKYSGODEVITGP
AAGRHSRDEGLITGFVNTLAIRKVSGRPTDPLHDKVTVIDAFHOYPERIA
EKVGASRDSLEIITFKDIAERMLKAYIHILEGRTADPLRIKGEYSILTEEPNQLQAE
GELLIDISTKFKKDIAERMLKAYIHILEGRTADPLRIKGEYSILTEEPNQLQAE
VOPASDTPREKTIYOLFEBOAERGHHALOEDKVMVSGELNRKNOLARLRGRTG
VSHRNLITVYNTYLDVYMSDLVLSISNYVDAEMFDEFGSILNKGALIAKDDIL
KAMKPLAAAEFGDILFEDEHLLAGASVYIPIDPPEKROHPEDEGAVLLQD
VMSRLAHLEKEKTYLMTITLFPNLIDMRDPSLKGIRVRLTSGEASDVADRRLN
TGKGRLLHMYGSESTVITTYVNEVDDOAPIGKPVSNLYLIDSDGNVPA
RVDOVKIRGORTELCEHNLVTHEMVOEAVALVGTGRKDMICVATTAODELFEIG
ELRYAAEGLPGMIFPSVMOLEPLTNGKNGRRLAPPEPDVQAQOAYVAPBSRP
GAOLADMOEVLNPRIGVHNFEFELGHSILGRTILARIQOEMWNYDOLDOAPR
IETLAQAEKTKSAVYIEAVPDROYVSSQKRLVYLOOIEGAKSYMPLAVDOL
EGKDLIKRLESAAOMLIRHEAFRTTEIYDGEVYVIMEAELITDVIIDDEBAG
LTIKERTRPDLKAPLEPMSTIRFTGKHLILVDMHIIISDASVSLIDMKIKLVG
GSRSLFALKPELVQOOLRKLAKETDSTLYVLAASVALLSGOSVITVDMKIKLVG
HTDLNRIIGMENVTLAIRPREDKPSAFLEVEKETLUGAFHODIYVGSAPARP
EMEYNTALFKHESIERRKGTWNLEAVENPDALSELILDAEKRRIRIYGARPEO
QDMKRNPLDFAVSMRNADIKOLSMEGITLKYDPAHOTAKPTIILTAEDDGLIVF
LAVPODTHLEPEAOYLTPDORGAAYVNDSPAPERTGYLDSGAELLITOTMLAA
RIGIMVRSLEMAAGVLAIIKGAAYVNDSPAPERTGYLDSGAELLITOTMLAA
PEEPSEGTILLDSMSEETTXDDEVNPOAGOSNDLAIYISGTTGQKGVYHRS
LVNLCWYHNDARVTEEDOKSAVAGFGDASVEMFPYUJAGAEHLIIIDEALRMDITR
LNOYPREKNTITTFPQLCEFMELNOSIRVLTGDKLRIKARSYTVLVNNGRT
ENTVATSTAIIDPKGMLSIGKPIANTRAYVLQONNEVOPVAGCELCIARGIARIG
LNPQARLESEVOEAVIDIDPAGCNKALCGIVANQDLEIGRISIDQVCIIRGRIEIS
FYVOLARLEPVANGVDRRALPODVBKJOTAEKAPLETBOLLADIMQELPDIWES
IDNNEFALGDSISIGIOMASNLQOHRKMLKMDLFOHPITIGELISAYVOQADDDIGRIG
PVSEGVTLTPIQWFERERRRTNBNHNSQVMJHALTGDPDETAKSLIIDEHDAI
EYARLEEVGSAFAKEITWQIELOAETALPRDDEAKRRHRIKAEELISIDETPL
TEOJMTVYHAYVNTBMNDLILALGALAKEMTGODPTICLGHGREDIIDGINSRT
VGFMTSOPPALIOMHNSDGIQVIOIKKEELRHPNGGIVGIRYLRITGEGKADPIK
HDSLEKSTVYAAASFYKYLQILDHICLERDGEELTPSDGLDELTLEELDKLMEITP
"
gene
11230.21996
/feature="1icB"
11230.21996
/feature="1icB"
/feature="1icB"
/function="antibiotic-biosynthesis"
/note="1icB"
/codon_start=1
/transl_table=1
```

```
/product="1ichenysin synthetase B"
/protein_id="AAD04758.1"
/db_xref="GI:3080743"
/translation="MGOKIKOVYVLETPMOGMLYHAMLDPESSYFTOLELFDGEF
DLGIFSEVOALVRSYDILITFYVHQOLQKROVVLAMERAIYERFDLADLDEEONN
RIDNYKQEOVAAQFNLAKMLPRTVFLDNRKTLVWVSHHILVWGMSWMLMRLEF
QNTAFANFANTVPLDQKRTADYIKMLGRQKGAASVYHNRILVWGMSWMLMRLEF
PDSERNEEFTVVDKRNVDLQOANRHYVAVPNIPOIMAVIAYLVNRTDVGK
TVVSGRSPSTDEHBAAGFINTIPVRLKYDONASRLEFKQAOHAIAIYARYVPL
VYGAQMTIKLTLNGAADAFAIIOAGHGLTIRIABEALINPSKASAMLSKOEOT
LIAFQNTASDIPREKTIPELEPEOVQASPKTRAAADGSLTYRIRKANKMOOET
VYKQSTOISLEGVEVLAVANMAEDKANLEVNKKEDELLAYIYISGSGTKGV
VHRNIVLVRNAGCIPKSGVMAOTAVSDASTFEVFGALLNGGTLIPVKEVLL
BSPELSLWNGCEPTEMTTSTCTLLDOAERTIDPKRIGNSTAYIDEGDIALPVKVR
DHOVYVGFPIELGEVEALALBESIISAAVIRENNAGNETCAIYAGARPAOLRT
ELSLSPEYVIMPAHILEDSMPLANGVDRQULPAIABETACIYAGARPAOLRT
IMBEILIKOPGIDNFTSISGHSIKAMMLAKIOEOLKENVIKVLEKPTIRLEAE
ERLEAERFALIRSHETLRTSELIDGETVOIHODVFIUTSVAEAEAAIARERT
RPELNRKAVIAELVLEKRLHILDMHIIITDSSGTLIRPLAGFYEAARABERT
KLHYKOPAWQOPEPHOKLEKROBYVIDVFKELDELSVGLIRPLAGFYEAARABERT
NIDGMVNTLALRNKPEGKTFPOLOVKEISLQGFARODYLPDFPRPKRSFEDRY
RPLDSVFNQNMNEVPLTKISDLOISSVHRRAKADLLEAVEHEDGICISFYA
VSLREESTEEKADMPPOPAVNGERTILYVRVLEKAOIANVIRSGVGRQAELE
LRSAAEAAALIVKMSGGAFLPIDMEPTLVRVLEKAOIANVIRSGVGRQAELE
ITGRLIDVRDITLMAKSNKRPHPTTGASIALIYISGTTGQCYOLERRNANVSW
FINDHLSDDKTMILSYAPDLGYSFVLLVAGELIYVRETIHEATHTIGE
OGITVYMLPFLSHTHYVOOSFALAKGLOSIRILVIGGERINKDERHSEYPTFR
INHPGPEITTIKAAPLEPDIOGFALOSIRILVIGGERINKDERHSEYPTFR
LYITGGLRAGVNNBELTAKFVANYPTGPTGSGALVLRNLNGSELVYAGAGE
VKIRGRIRIKETIEAFNGLDIERAVITFTGDSGLDELCAVIOAKQOLVSLRAC
LSERPDLVYKQGLDNEFDLGSILKGMWLAIEIOSRIGKGVPLKTLDPDETACIA
GIMPEYLVQKQGLDNEFDLGSILKGMWLAIEIOSRIGKGVPLKTLDPDETACIA
PDDBRKRAFADLVNRHESLRTAFLEIDGRTVOIHEKADIDKLYIHIREDDAEILN
EKLRTDSTALIRAEELISSRHLIDTHNIIIDAVSDILKLYIHIREDDAEILN
EKLRTDSTALIRAEELISSRHLIDTHNIIIDAVSDILKLYIHIREDDAEILN
VRFKASQKALIKRLAHVNTLNTVMAVNFLEISLQGDIVYVGPVOSFADPR
LKDPMGAFVNSLAKNHVNTLNTVMAVNFLEISLQGDIVYVGPVOSFADPR
MSRDLPEKNTIEDPEDKTTDLDGLITKPEI SHAARKDPLGAFEDHRIKLPAD
YATDLPEKNTIEDPEDKTTDLDGLITKPEI SHAARKDPLGAFEDHRIKLPAD
PEDVYHLEFEAOYLTPDORGAAYVNGVEMTYKELANRANLARIYONNETGLDY
IMKPSLEMAAGVLAIIKGAAYVNDSPAPERTGYLDSGAELLITOTMLAA
FTGDVIDILRESIIGDELYPEDDMNDPSORSDNLAIVLYISGTTGQKGVYHRS
LVNLCWYHNDARVTEEDOKSAVAGFGDASVEMFPYUJAGAEHLIIIDEALRMDITR
LNOYPREKNTITTFPQLCEFMELNOSIRVLTGDKLRIKARSYTVLVNNGRT
ENTVATSTAIIDPKGMLSIGKPIANTRAYVLQONNEVOPVAGCELCIARGIARIG
LNPQARLESEVOEAVIDIDPAGCNKALCGIVANQDLEIGRISIDQVCIIRGRIEIS
FYVOLARLEPVANGVDRRALPODVBKJOTAEKAPLETBOLLADIMQELPDIWES
IDNNEFALGDSISIGIOMASNLQOHRKMLKMDLFOHPITIGELISAYVOQADDDIGRIG
PVSEGVTLTPIQWFERERRRTNBNHNSQVMJHALTGDPDETAKSLIIDEHDAI
EYARLEEVGSAFAKEITWQIELOAETALPRDDEAKRRHRIKAEELISIDETPL
TEOJMTVYHAYVNTBMNDLILALGALAKEMTGODPTICLGHGREDIIDGINSRT
VGFMTSOPPALIOMHNSDGIQVIOIKKEELRHPNGGIVGIRYLRITGEGKADPIK
HDSLEKSTVYAAASFYKYLQILDHICLERDGEELTPSDGLDELTLEELDKLMEITP
"
gene
22033.25901
/feature="1icC"
22033.25901
/feature="1icC"
/feature="1icC"
/function="antibiotic-biosynthesis"
/note="1icC"
/codon_start=1
```

```
/tranyl_table=11
/product="Iichenysin synthetase C"
/protein_id="A004759.1"
/db_xref="GI:3080744"
/translation="MKGDYILSQKKKHVQDIYVLSPMQGMLEHTLHPGQSFYIEQ
ISMVQVGSFOKDIIEKSMNVYIGRYDIFRTVFEHKKRPQVVLKERSFOAEIDLS
GLSEAEONERIEDYKRRDKKGFNLKSDIPMTAVFKKGODREWWSYHHLLDGMG
FGIVQELFEVYNALRENRLYSGLPKYKDYIKWLESQDKORSILAVMEHHLAGEFGQ
TTPSEORKKPEQAGNQPPELLFTLPKEDTEVFTQAKAHHTTLSTALOAVSVLSRY
QSRDLIFGTVVSGRPADIOGVEHMGFLINVVPKRVTFGAHTTFTELVAELQKQALE
AEPHQYMPLYDIQSHIAPNLIDHIIVFENYPLQEBANKQHEEKNLGFTMDITVFEKS
NYDLNLASPGEEMLKLAVNRNVFEPSTILOKEQLTAIREIITNPEOPLDDICV
SKKEERLLHEENGVOVKOTGLTIPOFEQCAALPEQPAVSAAGRTLTYRELNEQA
NKVAHLRRKNRGEPVALLFRSPENWIAITAVLKAGAVLPIDPEYPEARIQYML
EDSGAVCMLTQELAGQAASLFRHNTILIDPAVSAESGRNLEIAEPDOLAVIMYT
```

```
Query Match 4.1%; Score 20; DB 1; Length 28798;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 229 CAGGATAAATAAAGCTAAA 248
|||||
Db 27381 CAGGATAAATAAAGCTAAA 27400
```

Search completed: October 27, 2002, 21:48:35  
Job time : 1558.38 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 : Search time 18.1865 seconds  
(without alignments)  
1540.986 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836  
Sequence: 1 MNRGCLQGSSLIITSVFLVNG.....ITGTACRQPDGRMQUISTFK 162

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_villus:\*  
16: sp.\_bacteriap:\*  
17: sp.\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	836	100.0	162	2	Q9F9K8
2	313.5	37.5	148	2	054381
3	291	34.8	159	2	Q9F9F2
4	266.5	31.9	137	2	052252
5	260.5	31.2	144	2	Q9K2N6
6	259.5	31.0	137	2	031065
7	259.5	31.0	144	2	Q9K4W8
8	252	30.1	154	2	053154
9	247	29.5	151	2	Q9F9G9
10	237.5	28.4	131	2	09F001
11	237.5	28.4	131	2	052637
12	236.5	28.3	131	2	09L522
13	223.5	26.7	105	2	031208
14	143	17.1	199	16	Q985G4
15	140	16.7	77	2	Q9AGC7
16	130.5	15.6	136	16	Q92R89

17	128.5	15.4	182	16	Q9HX13	Q9HX13 pseudomonas
18	116	13.9	154	16	Q9I4S1	Q9I4S1 pseudomonas
19	112	13.4	257	16	Q9PGX0	Q9PGX0 xyella fas
20	111	13.3	139	2	052854	052854 rhizobium 1
21	110.5	13.2	232	16	Q9A3X8	Q9A3X8 caulobacter
22	108.5	13.0	155	2	Q9F6B1	Q9F6B1 edwardsiella
23	107.5	12.9	153	2	069776	069776 rhizobium e
24	107.5	12.9	155	2	Q9RA95	Q9RA95 serratia sp
25	106.5	12.7	221	16	Q92ST9	Q92ST9 rhizobium m
26	106	12.7	155	2	Q9RB08	Q9RB08 pectobacter
27	105.5	12.6	223	2	Q9XCA4	Q9XCA4 porphyromon
28	101.5	12.1	304	16	Q9I762	Q9I762 pseudomonas
29	99.5	11.9	83	16	Q92LP2	Q92LP2 rhizobium m
30	99	11.8	154	16	Q9CN83	Q9CN83 pasteurilla
31	98	11.7	172	16	P76572	P76572 escherichia
32	98	11.7	220	2	Q9XAX8	Q9XAX8 pseudomonas
33	96	11.5	223	16	Q9KSR1	Q9KSR1 vibrio chol
34	95.5	11.4	79	16	Q92NT4	Q92NT4 shigella fl
35	94.5	11.3	838	2	Q9AL49	Q9AL49 arabisopsis
36	94	11.2	608	10	Q9SUX1	Q9SUX1 caulobacter
37	93.5	11.2	257	16	Q9A8W8	Q9A8W8 rhizobium 1
38	92	11.0	105	16	Q9BP93	Q9BP93 nephila cla
39	91	10.9	544	5	046171	046171 rhizobium 1
40	90	10.8	105	16	Q983Y0	Q983Y0 pseudomonas
41	89	10.6	137	16	Q9HU07	Q9HU07 bacillus me
42	88.5	10.6	161	2	068681	068681 rhizobium 1
43	88.5	10.6	216	16	Q98NA9	Q98NA9 bacillus su
44	87.5	10.5	407	16	Q31557	Q31557 rhizobium m
45	87	10.4	201	16	Q930X4	Q930X4 rhizobium m

## ALIGNMENTS

### RESULT 1

ID	Q9F9K8	PRELIMINARY:	PRT:	162 AA.
AC	Q9F9K8:			
DT	01-MAR-2001 (TREMUREL. 16, Created)			
DT	01-MAR-2001 (TREMUREL. 16, Last sequence update)			
DT	01-MAR-2001 (TREMUREL. 16, Last annotation update)			
DE	17 KDA ANTIGEN.			
GN	OSPA.			
OS	Piscirickettsia salmonis.			
OC	Bacteria; Proteobacteria; gamma subphylum; Piscirickettsia group;			
OC	Piscirickettsia.			
OX	NCBI_TaxID=1238;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-LF-89;			
RA	Kuzuk M.A., Burton J., Thornton J.C., Kay W.W.;			
RT	"Identification of a genus-common Rickettsial surface antigen in the			
RT	RT salmonid pathogen Piscirickettsia salmonis.";			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF184152; AAC17000.1; -			
SO	SEQUENCE 162 AA; 17661 MW; DDE99E6FD94A527E CRC64;			

Query Match	100.0%;	Score 836;	DB 2;	Length 162;
Best Local Similarity	100.0%;	Pred. No. 9.6e-68;		
Matches 162;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MNRGCLQGSSLIITSVFLVNGCAONFSGROEVAATGAVGAGGLPFGSGSRVMAATGA	60	
DB	1	MNRGCLQGSSLIITSVFLVNGCAONFSGROEVAATGAVGAGGLPFGSGSRVMAATGA	60	
QY	61	VLGLISKTIQSDMDQDKIRLNOSTLEKVRAGOVTRRNPDPTGSSVEVPRTQRYNKK	120	
DB	61	VLGLISKTIQSDMDQDKIRLNOSTLEKVRAGOVTRRNPDPTGSSVEVPRTQRYNKK	120	
QY	121	ERRQOYCRFQOKAMIGAKQOEIYGTACROPDGRMQUISTFK	162	
DB	121	ERRQOYCRFQOKAMIGAKQOEIYGTACROPDGRMQUISTFK	162	

```

RESULT 2
ID 054381
AC 054381; PRELIMINARY; PRT; 148 AA.
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DE 01-NOV-1998 (TREMBLrel. 06, Last sequence update)
OS 17 KDA COMMON-ANTIGEN (FRAGMENT).
OC Rickettsia sp.
Bacteria; Proteobacteria; Rickettsiales; Rickettsia.
NCBI_TaxID=789;
RP SEQUENCE FROM N.A.
RX MEDLINE=98087556; PubMed=9425244;
RT "Rickettsial relative associated with papaya bunchy top disease."
RL Curr. Microbiol. 36:80-84(1998).
DR EMBL, U76907; AAC02809.1; -.
FT NON_TER 1 148 1
SQ SEQUENCE 148 AA; 15050 MW; A7AFEPDEDAEEAC CRC64;

Query Match
Best Local Similarity 37.5%; Score 313.5; DB 2; Length 148;
Matches 57; Conservative 29; Mismatches 42; Indels 9; Gaps 2;

OY 26 SROEVGATGAVGVAGOLFSGSRVAMATGAVLGLISGKISQSDQDK----IK 81
DB 17 NKQSGTTLIGTLGLVSGFSGGGRILAAGAGALGALGAGNDQDDEKRLAELT 76
OY 82 LNSLEKRVAGVTRWRNPDGNSYSVEPYRTQRTNKRQOYCRFFQOKAMIAOKO 141
DB 77 SORALPAAPSGSSVQMRNPDNGNYGVTPSKAY-----KNNTGYCRBYTQTVVGSKQ 131
OY 142 EYGTACRQPPGQROVY 158
DB 132 KAVGTACRQPPGQROVY 148

RESULT 3
OY9F9F2
AC 09F9F2; PRELIMINARY; PRT; 159 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DE 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
OS 17 KDA GENUS-COMMON ANTIGEN.
OC Rickettsia felis (Rickettsia azadi).
Bacteria; Proteobacteria; Rickettsiales; Rickettsia.
NCBI_TaxID=42862;
RP SEQUENCE FROM N.A.
RX MEDLINE=21217364; PubMed=11321078;
RT Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,
RA Zavala-Velazquez J.E., Foil L.D., Storchard D.R., Azad A.F.,
RT "Rickettsia felis: molecular characterization of a new member of the
RL spotted fever group."
DR EMBL, AF195118; AAC28452.1; -.
FT NON_TER 1 159 1
SQ SEQUENCE 159 AA; 16497 MW; 34C5B020FA70A1F CRC64;

Query Match
Best Local Similarity 34.8%; Score 291; DB 2; Length 159;
Matches 61; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

OY 9 SSLLITSV--FLVGC--AQNFSROEVGATGAVGVAGOLFSGSRVAMATGAVLG 63
DB 5 SKMIALAASMLQACNGPGGMKNGGTGTLTGAGGALLGSGFGKGGQLVGVGALLG 64

```

```

OY 64 GLISKISGSDQDK-----ILNOSLEKRVAGVTRWRNPDGNSYSVEPYRTQRTN 119
DB 65 AVIGGQIGACMDQDRLAELTQSRALATPSSGVEMRNPNGNHYVTNKT----- 119
OY 120 QERRQOYCRFFQOKAMIAOKOYGTACRQPPGQROVYS 159
DB 120 RNSTGYCRBYTQTVVGSKQKAVGACRQPPGQROVYN 159

RESULT 4
ID 052252
AC 052252; PRELIMINARY; PRT; 137 AA.
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DE 01-DEC-2001 (TREMBLrel. 06, Last sequence update)
OS 17 KDA ANTIGEN (FRAGMENT).
OC Rickettsia coolieyi.
Bacteria; Proteobacteria; Rickettsiales; Rickettsia.
NCBI_TaxID=69410;
RP SEQUENCE FROM N.A.
RX Billings A.N., Tellow G.J., Walker D.H.;
RT "Molecular characterization of a novel spotted fever group rickettsial
RL species from Ixodes scapularis in Texas."
DR EMBL, AF031534; AB95267.1; -.
FT NON_TER 1 137 1
SQ SEQUENCE 137 AA; 14215 MW; A27597A9AFDB5FC3 CRC64;

Query Match
Best Local Similarity 31.9%; Score 266.5; DB 2; Length 137;
Matches 53; Conservative 23; Mismatches 49; Indels 9; Gaps 2;

OY 23 ONFSROEVGATGAVGVAGOLFSGSRVAMATGAVLGLISGKISQSDQDK--- 79
DB 7 KGNKQGTGTLTGAGGALLGSGFSGGGRILAAGAGALGAGNDQDDEKRLAELT 66
OY 80 -ILNOSLEKRVAGVTRWRNPDGNSYSVEPYRTQRTNKRQOYCRFFQOKAMIAOKO 138
DB 67 ELTQSRALPAAPSGSSVQMRNPDNGNYGVTPNKT-----RNSTGYCRBYTQTVVG 121
OY 139 OKOELGTACRQPP 152
DB 122 KQKAVGACRQPP 135

RESULT 5
OY9K2N6
AC 09K2N6; PRELIMINARY; PRT; 144 AA.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DE 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
OS 17 KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).
OC male-killing Rickettsia from Adalia bipunctata.
Bacteria; Proteobacteria; Rickettsiales; Rickettsia.
NCBI_TaxID=38028;
RP SEQUENCE FROM N.A.
RX Schuenburg H.-J.G.V.D., Habig M., Sloggett J.J., Webberley M.R.,
RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;
RT "On the evolution of male-killing: Monophyletic origin and horizontal
RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-
RL (Coleoptera: Ladybirds, Adalia bipunctata L. and A. decempunctata L.
DR EMBL, AJ269518; CAB96383.1; -.
FT NON_TER 1 144 1
SQ SEQUENCE 144 AA; 144 1

```

```

DE 17 KDA ANTIGEN (FRAGMENT).
OC male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=120393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20575219; PubMed=11133455;
RA Schuenburg H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot Ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT Coccinellidae).";
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL; AJ269516; CAB96381.1; -.
FT NON_TER 1 144
FT NON_TER 1 144
SO SEQUENCE 144 AA; 14801 MW; C825472F16A56AE7 CRC64;

Query Match 31.0%; Score 259.5; DB 2; Length 144;
Best Local Similarity 39.7%; Pred. No. 6,3e-16;
Matches 52; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SRQEVGATGAVGVGAQQLFGKSGRRVAMATGAVLGLGKSGKIGQSMDOODK-----IK 81
DB 17 NKQGTGTLGGAGGALLDSQFGKGGQLVGVGVALLAVALGQIGAAMDQDRRLAELT 76
QY 82 LNSLEKVKAGVTRMRNPDTGNSYSVEPVRTYGRYRKNQESRQOYCFREFQOKAMATAGOK 141
DB 77 SQRALEAPSGSNVEMRNPDMNGHGYVPNKTY-----RNSIQYCREYTGTVVIGCKQO 131
QY 142 EITGTACROPD 152
DB 132 KSYGNACROPD 142

RESULT 8
Q53154 PRELIMINARY; PRT; 154 AA.
AC Q53154;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, last annotation update)
DE (CLONE PRB F15F 1), 5' END CDS (FRAGMENT).
OS Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=789;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93084757; PubMed=1452660;
RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
RT "Characterization and comparison of Australian human spotted fever
RT group Rickettsiae."
RL J. Clin. Microbiol. 30:2896-2902(1992).
DR EMBL; M99391; AAA73386.1; -.
FT NON_TER 154
FT NON_TER 154
SO SEQUENCE 154 AA; 15849 MW; F5C35855EDBA39D2 CRC64;

Query Match 30.1%; Score 252; DB 2; Length 154;
Best Local Similarity 35.5%; Pred. No. 3,3e-15;
Matches 55; Conservative 29; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIIV---FLVCC--AQNFSROEVGATGAVGVGAQQLFGKSGRRVAMATGAVL 63
DB 5 SKIMIIATATSLWACNCPGCMKNGKGTTLGGAGGALLGDSFGKGGKQOLVGVGALLG 64
QY 64 GLISKIGQSMDOODK---IKLNSLEKVKAGVTRMRNPDTGNSYSVEPVRTYQRYNK 119
DB 65 AVLGQIGAGMDEQDRRLAELTSQLALETAPSGSNVEMRNPDMNGHGYVPNKTYRNSTG 124
QY 120 QERRQYCFREFQOKAMATAGOKEITGTACROPD 154

```

Db 125 QD-----CRYVTVTVIGKQKAYGNACRQPDGQ 154

## RESULT 9

Q9F909 PRELIMINARY; PRT; 151 AA.  
 AC Q9F909;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE OUTER MEMBRANE PROTEIN (FRAGMENT).  
 OS Rickettsia helvetica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN NCBI\_TaxID=35789;  
 RP SEQUENCE FROM N.A.  
 RA Nilsson K., Pahlson C.;  
 RT "Novel peptide diagnostic reagent and kit for detection of  
 Rickettsiosis."  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL AF181036; MAG09427.1;  
 FT NON\_TER 151  
 SQ SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match  
 Best Local Similarity 29.5%; Score 247; DB 2; Length 151;  
 Matches 55; Conservative 26; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIIV---FLVGC--AQNFSRQEVGATGAVGVAGOLFGRSGRVAMAIGGAVIG 63  
 Db 5 SKIMIIAASMLQACNGPGMNKQGTGILGAGAGALLGSQFGKQGLVGVGVALLG 64  
 QY 64 GLISKTIQSDQDQK---IKLNSLEKRVKAGQVTRMRNPDTGSSVSEVRYQRYNK 119  
 Db 65 AVLGQIVAGMDQDRLRLLELTQSRLAEPSSNVEMRPNNGVYVTPNKTY----- 119  
 QY 120 QERRQYCRFEQOKAMIAQOKQETIGTACROP 151  
 Db 120 RNSTGYCREYTVTVIGKQKAYGNACRQPDGQ 151

## RESULT 10

Q9F001 PRELIMINARY; PRT; 131 AA.  
 AC Q9F001;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 17 KDA PROTEIN (FRAGMENT).  
 OS Rickettsia sp. California 2.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN NCBI\_TaxID=147259;  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CALIFORNIA 2;  
 RT Raoult D.;  
 RL "A new SFG rickettsia isolated from fleas."  
 RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Roux V., Raoult D.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL AF210693; AAC48554.1;  
 FT NON\_TER 131  
 SQ SEQUENCE 131 AA; 13374 MW; 2308819B29FF860 CRC64;

Query Match  
 Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;  
 Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGVAGOLFGRSGRVAMAIGGAVIGLIGSKTIQSDQDQK----IK 81  
 Db 10 NKQGTGILGAGAGALLGSQFGKQGLVGVGVALLGAGIGAGMDQDRLRLLELT 69  
 QY 82 LNSLEKRVKAGQVTRMRNPDTGSSVSEVRYQRYNKQERRQYCRFEQOKAMIAQOK 141  
 Db 70 SORALEAPSSNVEMRPNNGVYVTPNKTY-----RNSTGYCREYTVTVIGKQ 124  
 QY 142 EITGTAC 148  
 Db 125 KAYGNAC 131

## RESULT 11

Q52637 PRELIMINARY; PRT; 131 AA.  
 AC Q52637;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 17 KDA ANTIGEN (FRAGMENT).  
 OS Rickettsia sp.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN NCBI\_TaxID=789;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94117373; PubMed=8288533;  
 RA Warren J.H., Hurst G.D., Zhang W., Breeuwer J.A., Stouthamer R.,  
 RT "Rickettsial relative associated with male killing in the ladybird  
 beetle (Adalia bipunctata)."  
 RL J. Bacteriol. 176:388-394(1994).  
 DR EMBL: U04162; AAA19233.1;  
 FT NON\_TER 131  
 SQ SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

Query Match  
 Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;  
 Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGVAGOLFGRSGRVAMAIGGAVIGLIGSKTIQSDQDQK----IK 81  
 Db 10 NKQGTGILGAGAGALLGSQFGKQGLVGVGVALLGAGIGAGMDQDRLRLLELT 69  
 QY 82 LNSLEKRVKAGQVTRMRNPDTGSSVSEVRYQRYNKQERRQYCRFEQOKAMIAQOK 141  
 Db 70 SORALEAPSSNVEMRPNNGVYVTPNKTY-----RNSTGYCREYTVTVIGKQ 124  
 QY 142 EITGTAC 148  
 Db 125 KAYGNAC 131

## RESULT 12

Q9L522 PRELIMINARY; PRT; 131 AA.  
 AC Q9L522;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 17 KDA SURFACE ANTIGEN (FRAGMENT).  
 OS Rickettsia peacockii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN NCBI\_TaxID=47589;  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-DAE100R;  
 RX MEDLINE=21091941; PubMed=11157215;  
 RX Sinsler J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 5.87565 Seconds  
(without alignments)  
1067.553 Million cell updates/sec

Title: US-09-677-374-2

Sequence: 1 MNRGCLQGSSLIITISVFLV.....IYGTACRQPDGRMNVISTEK 162

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296	35.4	159	17KD_RICPR	P16624 rickettsia
2	294	35.2	159	17KD_RICJA	Q52764 rickettsia
3	293	35.0	159	17KD_RICCN	P05372 rickettsia
4	284.5	34.0	159	17KD_RICCT	P22882 rickettsia
5	276	33.0	154	17KD_RICAU	P60928 rickettsia
6	263	31.5	154	17KD_RICAM	P50927 rickettsia
7	261	31.2	154	17KD_RICPA	P50930 rickettsia
8	261	31.2	154	17KD_RICRM	P50931 rickettsia
9	258	30.9	154	17KD_RICMO	P50929 rickettsia
10	157	18.8	80	17KD_RICCA	P29687 rickettsia
11	113.5	13.6	155	PCP_YEREN	P1484 versinia en
12	105.5	12.6	155	SLVB_ECOLI	Q53549 escherichia
13	105.5	12.6	155	SLVB_SALTY	Q53549 haemophilus
14	102	11.8	155	PCP_HAERIN	P37376 escherichia
15	99	11.8	179	KCFJ_ECOLI	P06394 bos taurus
16	90	10.8	526	KICJ_BOVIN	P4334 bacillus su
17	90	10.8	1332	XKDO_BACSU	P45931 bacillus su
18	89	10.6	1585	YOBQ_BACSU	P37723 salmonella
19	87.5	10.5	72	OSMB_SALTY	P17853 salmonella
20	86.5	10.3	72	OSMB_ECOLI	P17853 escherichia
21	84.5	10.1	243	CYSH_SALTY	P18837 rickettsia
22	84.5	10.1	301	STXG_RAT	Q00310 rattus norv
23	84.5	10.1	431	KRE2_CANAL	Q00310 candida alb
24	84.5	10.1	526	VPS_BTUVI	P33476 blueconque
25	82	9.8	132	V615_AQVAE	O66867 aquilex aeo
26	82	9.8	263	CANS_BOVIN	P13135 bos taurus
27	81.5	9.7	541	NU57_YEAST	P18837 saccharomyc
28	81.5	9.7	747	SPD1_MERCL	P18837 nephila cia
29	80.5	9.6	467	HEM1_MYCLE	P46724 mycobacteri
30	79.5	9.5	113	YKR3_CAEEL	P34309 cryocollagus
31	79.5	9.5	266	CANS_RABIT	P06813 onychomys
32	79.5	9.5	359	ATPA_BOVIN	P19482 bos taurus
33	79.5	9.5	543	ATPA_RAT	P15999 rattus norv

## ALIGNMENTS

34	79.5	9.5	553	1	ATPA_HUMAN	P25705 homo sapien
35	79	9.4	593	1	KICJ_HUMAN	P13645 homo sapien
36	78.5	9.4	553	1	ATPO_BOVIN	P03483 bos taurus
37	78.5	9.4	553	1	ATPA_MOUSE	Q03263 mus musculus
38	78	9.3	865	1	VGLB_HSVMD	P18538 marek's dis
39	77.5	9.3	219	1	YIAD_ECOLI	P37665 escherichia
40	77	9.2	269	1	CANS_MOUSE	O88456 mus musculu
41	77	9.2	393	1	CSP_PLABR	P14593 plasmodium
42	77	9.2	429	1	CSP_PLAMA	P13815 plasmodium
43	77	9.2	506	1	ATPA_ANASP	P12405 anabena sp
44	77	9.2	514	1	ATPA_THIFE	P1167 thiodacillu
45	76.5	9.2	243	1	CYSH_ECOLI	P17854 escherichia

  

RESULT 1	17KD_RICPR	STANDARD:	PRT:	159 AA.
AC	P16624;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	17 kDa surface antigen precursor.			
GN	OMP OR RP833.			
OS	Rickettsia prowazekii.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Rickettsiae; Rickettsia.			
OX	NCBI_Taxid=782;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MADRID E;			
RC	MEDLINE=89359171; PubMed=2768201;			
RA	Anderson B.E., Tzianabos T.;			
RT	"Comparative sequence analysis of a genus-common rickettsial antigen gene";			
RL	J. Bacteriol. 171:5199-5201(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MADRID E;			
RC	MEDLINE=99039499; PubMed=9823893;			
RA	Anderson S.G.E., Zomrodipour A., Andersson J.O.,			
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,			
RT	Eriksson A.-S., Winkler H.H., Kurland C.G.;			
RT	"The genome sequence of Rickettsia prowazekii and the origin of mitochondria.";			
RL	Nature 396:133-140(1998).			
CC	-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M28482; AAA26378.1; ALT_SEQ.			
DR	EMBL: AJ235273; CAA15258.1; -			
DR	PIR: D33971; D33971.			
DR	PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.			
KW	Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.			
FT	SIGNAL	1	19	
FT	CHAIN	20	159	17 KDA SURFACE ANTIGEN.
FT	LIPID	20	20	N-ACYL DIGLYCERIDE (PROBABLE).
SQ	SEQUENCE	159 AA:	16672 MW;	A33D404B5EBB071 CRC64;

  

Query Match	35.4%	Score 296;	DB 1;	Length 159;
Best Local Similarity	38.1%	Pred. No. 9.9e-19;		
Matches	61;	Conservative	30;	Mismatches 55; Indels 14; Gaps 4;

Qy 9 SLLIIIV--FLVGC--AQNFSOEVGATGAVVGAGVGLGCKSGGVAMALGGAVLG 63  
 Db 5 SKMIIITALLAAMLCACNCGSNMKGSTGLLGAGGALLGCSQGGKQKQLGVGVGALLG 64  
 Qy 64 GLIGSKTIGOSMPODK----IKTNOSLEKVGAVTPFNRBDTGSYSEFVRYORYNK 11  
 Db 65 AVTGGQIASMDQEDDRRLLELISORALSSAPSGSNTLEMRNPNGHGVTPNPKY---- 11  
 Qy 120 QERRQGYCFREFQOKAMIAQCKEYITGACNPGGRNOVYS 159  
 Db 120 RNSAGQYCFREYTOTVLISGRQDKTYGNACRDPGQWQVYN 159

RESULT 2	17KD.RICJA	STANDARD;	PRT;	159 AA.
ID	17KD.RICJA			
AC	052764;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DE	17 Kda surface antigen precursor.			
GN	OMP.			
OS	Rickettsia japonica.			
OC	Bacteria; Proteobae.			
OC	Rickettsiaceae; Rickettsiales; Rickettsia.			
OX	NCBI_TaxID=35790;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-YH.			
RX	MEDLINE=95229950; PubMed-7714214;			
RA	Furuya Y., Katayama T., Yoshida Y., Kaiho I.;			
RT	"Specific amplification of Rickettsia japonica DNA from clinical			
RT	specimens by PCR.";			
RL	J Clin. Microbiol. 33:487-489(1995).			
CC	-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid			
CC	anchor (Probable).			
CC	-----			
CC	This SWISS-PROT entry is Copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).			
CC	-----			
DR	EMBL; D1513; BAA03965.1; -			
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.			
KW	Outer membrane; Lipoprotein; Antigen; Signal.			
FT	SIGNAL			
FT	CHAIN			
FT	LIPID			
50	SEQUENCE			
159 AA;	16554 MW;			
N ACTV	DIGLYCERIDE (BPOH-1)			

```

Query Match      35.2% Score 294 DB 1 Length 159;
Best Local Similarity 38.1% Pred. No. 1.5e-18;
Matches 61; Conservative 30; Mismatches 55; Indels 14; Gaps
Qy          9 SLLIISV---FLVGC--AQNFSRQEVGATGVAVGGVAGACGLEFGSGRGVAAIGAVILG 63
Db          5 SKIMIIATLSMLACNCPGMKKOCTGTILLGAGAGALLSOFKGTGLQVLGVGVGALLG 64
Qy         64 GLIGSYIGSDMOODK-----IKINSLIEYVAKGVTVWRNPDTGNSYSVEPYRYQRNK 119
Db         65 AVLGGDIACGHDEDDRLAEELTSORALETAIAGSSNNEMRRPDPNGNVGYVTYNKTY 119
Qy        120 QERRQQCRREQQKAMIAQOKOEIYGTGAQPPGRMOVIS 159
Db        120 RNSTGQTGREYTOTVVIGGKOOKAKVGNACKAPDGOVOVYN 159

RESULT 3
ID_17KD_RICCN ID_17KD_RICCN STANDARD; PRT; 159 AA.

```

AC 005372.  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
CN 17 kDa surface antigen precursor.  
OS Rickettsia conorii,  
OS Rickettsia conorii, and  
OC Bacteria; Proteobacterales;  
OC Rickettsiales; Rickettsiaceae;  
OX NCBI\_TaxID=781, 783.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=R.conorii, and R.rickettsii;  
RX MEDLINE=69359171; PubMed=2768201;  
RA Anderson B.E., Tzianabos T.;  
RT "Comparative sequence analysis of a genus-common rickettsial antigen gene";  
RL J Bacteriol. 171:5199-5201(1989).  
RN [2]  
RP Bacteriol. 171:5199-5201(1989).  
RC SPECIES=R.conorii.  
RX MEDLINE=21442074; PubMed=3139629;  
RA Ogata H., Audic S., Kenealy-Audiffren P., Fournier P.-E., Barde V.,  
RT Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RL Raoult D.;  
RN [3]  
RP "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."; Science 293:2093-2098(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=R.rickettsii;  
RX MEDLINE=87222152; PubMed=3108232;  
RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McNamee J.E.,  
RT Fu Z.X., Bellini W.J.;  
RL "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia rickettsii.";  
RN J. Bacteriol. 169:2385-2390(1987).  
RP SEQUENCE OF 1-30 FROM N.A.  
RC SPECIES=R.rickettsii;  
RX MEDLINE=89008059; PubMed=3139629;  
RA Anderson B.E., Baumstark B.R., Bellini W.J.;  
RT "Expression of the gene encoding the 17-kilodalton antigen from Rickettsia rickettsii: transcription and posttranslational modification.";  
RN J. Bacteriol. 170:4493-4500(1988).  
RP -I-SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL; M28479; AA226379.1;  
DR EMBL; M28480; AAA26376.1;  
DR EMBL; AE008675; AAL03825.1;  
DR EMBL; M16486; AAA26381.1;  
DR EMBL; J03371; NOT\_ANNOTATED\_CDS.  
DR PIR; A25972; A25972.  
DR PIR; A31836; A31836.  
DR PIR; A33971; A33971.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.  
FT CHAIN 1 19  
FT LIPID 20 159  
FT FT 20 20  
FT FT 146 146  
FT CONFLICT 153 153  
G -> E (IN REF. 3)



RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U11013; AAB07704.1; -  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20  
 FT NON\_TER 154 154 N-ACYL DIGLYCERIDE (PROBABLE).  
 SQ SEQUENCE 154 AA; 15879 MW; E4FBEC29D943581 CRC64;

Query Match  
 Best Local Similarity 31.5%; Score 263; DB 1; Length 154;  
 Matches 57; Conservative 27; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIISV---FLVGC--AONFSROEVAATGAAGVAGOLFGKSGRYAMAGAVLG 63  
 DB 5 SKIMTIALAASMLQACNGPGMKGOTGTLGAGGALLGSGFGKGLVGVGALLG 64  
 QY 64 GLIGSKIGQSMDOOK-----IKLNSLEKVKAGVTRRRNDPTGNSVSEPVRTQRYNK 119  
 DB 65 AVLGQIGAGNDEDDRIAEITLSQALETPAPNSVEMRNPDNGNYGYPNKTY----- 119  
 QY 120 QERRQYCREFOQKAMTAGOKETIYGTACROPDGR 154  
 DB 120 RNSGQYCREYTOYVIGGKQAKAYGNACROPDQ 154

RESULT 7  
 17KD\_RICPA STANDARD; PRT; 154 AA.  
 AC P50930;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia parkeri.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsieae; Rickettsia.  
 RN NCBI\_TaxID=35792;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MACULATUM;  
 RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL: U17008; AAA82040.1; -  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT NON\_TER 154 154  
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45FDDBD5EEC CRC64;

Query Match  
 Best Local Similarity 31.2%; Score 261; DB 1; Length 154;  
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIISV---FLVGC--AONFSROEVAATGAAGVAGOLFGKSGRYAMAGAVLG 63  
 DB 5 SKIMTIALAASMLQACNGPGMKGOTGTLGAGGALLGSGFGKGLVGVGALLG 64  
 QY 64 GLIGSKIGQSMDOOK-----IKLNSLEKVKAGVTRRRNDPTGNSVSEPVRTQRYNK 119  
 DB 65 AVLGQIGAGNDEDDRIAEITLSQALETPAPNSVEMRNPDNGNYGYPNKTY----- 119  
 QY 120 QERRQYCREFOQKAMTAGOKETIYGTACROPDGR 154  
 DB 120 RNSGQYCREYTOYVIGGKQAKAYGNACROPDQ 154

RESULT 8  
 17KD\_RICRH STANDARD; PRT; 154 AA.  
 AC P50931;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia rhipicephali.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsieae; Rickettsia.  
 RN NCBI\_TaxID=33992;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL: U11020; AAB07706.1; -  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20  
 FT NON\_TER 154 154 N-ACYL DIGLYCERIDE (PROBABLE).  
 SQ SEQUENCE 154 AA; 15895 MW; 0CF85AD5D96DEFFB CRC64;

Query Match  
 Best Local Similarity 31.2%; Score 261; DB 1; Length 154;  
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIISV---FLVGC--AONFSROEVAATGAAGVAGOLFGKSGRYAMAGAVLG 63  
 DB 5 SKIMTIALAASMLQACNGPGMKGOTGTLGAGGALLGSGFGKGLVGVGALLG 64  
 QY 64 GLIGSKIGQSMDOOK-----IKLNSLEKVKAGVTRRRNDPTGNSVSEPVRTQRYNK 119  
 DB 65 AVLGQIGAGNDEDDRIAEITLSQALETPAPNSVEMRNPDNGNYGYPNKTY----- 119  
 QY 120 QERRQYCREFOQKAMTAGOKETIYGTACROPDGR 154  
 DB 120 RNSGQYCREYTOYVIGGKQAKAYGNACROPDQ 154

```

RESULT 9
17KD_RICMO STANDARD; PRT; 154 AA.
ID 17KD_RICMO
AC P50929;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OHIO 83-441;
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: U11017; AAB07705.1; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E31DA CRC64;

Query Match 30.9%; Score 258; DB 1; Length 154;
Best Local Similarity 36.1%; Pred. No. 1.7e-15;
Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIIV---FLVGC---AQNFSRQEVGAATGAVGVAGQLPEKSGSRVAMAGVAVL 63
DB 5 SKIITIALASMLQACPGCMNKGCTLLGAGALLSOFQGGQLVGVGVALLG 64
QY 64 GLISKIGSGMDQDK-----IKLNSLEKYAGOVTRMRNPDGNSVSEPVRTYQRYNK 119
DB 65 AVLOGQISGAGMDEDRRLALETISQRALETAPSGSNVEMRNPDNGNYGVTPNKTY----- 119
QY 120 QERRQGYCREFOOKAMITAGOKDEITYGTACROPDGR 154
DB 120 RNSTGYCREYGTQTVVIGKQOKAYGNACLOPDGO 154

RESULT 10
17KD_RICCA STANDARD; PRT; 80 AA.
ID 17KD_RICCA
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=92108069; PubMed=1729713;

```

```

RA Azad A.F., Sacci J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas."
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: M82879; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR000437; Prok_Lipoprot.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KM Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 157; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. NO. 4e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 43 GOLFSGSGRYAMAGVAVLGLIGSKIGSGMDQDK-----IKLNSLEKYAGOVTRMR 98
DB 1 GSGFGKGGQGLIGVAGALLGAILGNIGAGMDQDRRLALETISQRALETTPSTSIEMR 60
QY 99 NPDTGNSVSEPVRTYQ 115
DB 61 NPDGNGYGVTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD; PRT; 155 AA.
ID PCP_YEREN
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcv precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RL MEDLINE=92121089; PubMed=1731292;
RA Baunler A.J., Hanke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli."
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -----
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURUM SLXB AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: X60448; CAA42977.1; -

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX      MEDLINE=21074935; PubMed=11206551;
RA      Perna N.T., Plunkett G., Ill, Burland V., Mau B., Glasner J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA      Postal G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA      Grobeck E.J., Davis N.W., Lim A., Djmalanta E.T., Potamouls K.,
RA      Apoloca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RT      "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.",
RL      Nature 409:329-333(2001).
RN      [5]
RP      SOURCE FROM N.A.
RC      STRAIN-O157:H7 / RIND 0509952;
RX      MEDLINE=21156231; PubMed=11258786;
RA      Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA      Iida T., Takami H., Honda T., Sasaki M., Ogasawara N., Yasunaga T.,
RA      Kihara S., Shiba T., Hattori M., Shigenaga H.;
RT      "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12."
RL      DNA Res. 8:11-22(2001).
CC      -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      anchor (potential).
CC      -1- SIMILARITY: TO S. TYPHIMURITUM SLVB, H. INFLUENZAE PCP AND
CC      Y. ENTEROCOLITITICA PCP.
CC      -----
CC      CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      CC      the European Bioinformatics Institute. There are no restrictions on its
CC      CC      use by non-profit institutions as long as its content is in no way
CC      CC      modified and this statement is not removed. Usage by and for commercial
CC      CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      CC      or send an email to license@isb-sdb.ch).
DR      EMBL; AE000259; AAC74713.1; .
DR      EMBL; D90807; BAA15402.1; .
DR      EMBL; AE005387; AAG56630.1; .
DR      EMBL; AP002558; BAB35773.1; .
DR      EcoGene; EG13409; SLVB.
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Outer membrane; Lipoprotein; Signal.
FT      SIGNAL.
FT      CHAIN.
FT      LIPID.
FT      CONFLICT.
SQ      SEQUENCE 155 AA; 15602 MW; 543EBBA4069A5F83 CRC64;

Query Match
Best Local similarity 12.6%; Score 105.5; DB 1; Length 155;
Matches 33; Conservative 21; Mismatches 42; Indels 49; Gaps 4;

OY      13 IISFELVCAQN-----FSRQ-----VGAATGAV 37
DB      10 MWGLSLVGCANNDDTSGDYVTASFAKOVNVSGTIVNRPVQIGCGDSDNVATIGAV 69
OY      38 VGVAVAGQLRGKSGSGVVAATIGAVLGGILGSKIGSGMDQDQKIKNSLEKYNAGVTRW 97
DB      70 LGGFLGNTVGGGTGSLATAGAVAGVAGGCGVGSAMKKTGQVEL--EIRK----- 118
OY      98 RNPDDTNSYSEYEPRTYRYNKOER 122
DB      119 ---DDGNTINMYOKNGTREFSGOR 140

RESULT 13
ID      SLVB_SLATY
ID      SLVB_SLATY
AC      035349; STANDARD. PRT; 155 AA

```



GN SLVB OR STM1445 OR STY1677.  
OS Salmonella typhimurium, and  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_Taxid=602, 601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium;  
RX MEDLINE=96133688; PubMed=8544813;  
RA Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,  
RA Goebel W.;  
RT "SLVA, a regulatory protein from Salmonella typhimurium, induces a  
RT hemolytic and pore-forming protein in Escherichia coli.";  
RL Mol. Gen. Genet. 249:474-486(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhi; STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagsels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
CC anchor (Potential).  
CC -1- SIMILARITY: TO E.COLI SLVB, H.INFLUENZAE PCP AND Y. ENTEROCOLITICA  
CC PCP.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: S80790: AAB35871.2: -;  
CC DR EMBL: AE008762: AAL20367.1: -;  
CC DR EMBL: AL627271: CAD01922.1: -;  
CC DR StGene: SG10573: slvb.  
CC PROSITE: PS00013: PROKAR\_LIPOPROTEIN; 1.  
CC KM Outer membrane; Lipoprotein; signal; Complete proteome.  
CC FT STGNL 1 17  
CC FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLVB.  
CC FT LIPID 18 18 N-ACYL DIGLUCERIDE.  
CC FT SEQUENCE 155 AA: 15548 MW: 82FDCCBADD55A7 CRC64;  
CC  
CC Query Match 12.6%; Score 105.5; DB 1; Length 155;  
CC Best Local Similarity 24.6%; Pred. No. 0.02;  
CC Matches 34; Conservative 19; Mismatches 52; Indels 33; Gaps 3;  
CC  
CC 4 GCLQGSSL-----IIISFLVGCANFSGRQVGAATGAVGVAGQ 44  
CC 11: 11 1:1 1 11:11 1  
CC 17 GCVNNDLSGSDVYTAESAQVQVNTYGTIVNRPVQIOGGDDSNVIGALIGAVLVGFLGN 76  
CC DB

OY 45 LFGKSGRVAMALIGAVLIGLSKIGOSMDQDKIKLNSLEKVKAGQVTRMRPDTGN 104  
DB 77 TIGGCTGRSLATAGAVAGAGGQVQSGAMKRGVEL--EIRK-----DDGN 122  
OY 105 SYSEVPRTYQRYNKR 122  
DB 123 TIMVQKQGNTRFSAGOR 140  
DB  
DB RESULT 14  
DB PCP\_HAEIN  
DB ID PCP\_HAEIN STANDARD; PRT; 155 AA.  
DB AC P10325;  
DB DT 01-MAR-1989 (Rel. 10, Created)  
DB DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DB DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DB DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL  
DB DE cross-reacting lipoprotein).  
DB GN PCP OR LPP OR H1579.  
DB OS Haemophilus influenzae.  
DB OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
DB OC Haemophilus.  
DB OX NCBI\_Taxid=727;  
DB RN [1]  
DB RP SEQUENCE FROM N.A.  
DB RX MEDLINE=86115138; PubMed=2828309;  
DB RA Deich R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;  
DB RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated  
DB RT outer membrane lipoprotein and an antigenically related 15,000-dalton  
DB RT protein from Haemophilus influenzae.";  
DB RL J. Bacteriol. 170:489-496(1988).  
DB RN [2]  
DB RP SEQUENCE FROM N.A.  
DB RC STRAIN=RD / KW20 / ATCC 51907;  
DB RX MEDLINE=95350630; PubMed=7542800;  
DB RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
DB RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
DB RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
DB RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
DB RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
DB RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
DB RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
DB RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
DB RA Venter J.C.;  
DB RT "Whole-genome random sequencing and assembly of Haemophilus  
DB RT influenzae Rd.";  
DB RL Science 269:496-512(1995).  
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
CC anchor.  
CC -1- SIMILARITY: TO E.COLI AND S.TYPHIMURIM SLVB AND TO  
CC Y. ENTEROCOLITICA PCP.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M18877: AAA24938.1: -;  
CC DR EMBL: U32832: AAC23228.1: -;  
CC DR PIR: B28543; B28543.  
CC DR TIGR: H11579; -;  
CC PROSITE: PS00013: PROKAR\_LIPOPROTEIN; 1.  
CC KM Outer membrane; Lipoprotein; signal; Complete proteome.  
CC FT STGNL 1 18  
CC FT CHAIN 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.  
CC FT LIPID 19 19 N-ACYL DIGLUCERIDE.  
CC FT CONFLICT 135 143 CSLVAEFLV -> VAGRRVRI (IN REF. 1).  
CC FT SEQUENCE 155 AA: 15425 MW: D7880327FCFC985 CRC64;  
CC DB

```

Query Match          12.2%; Score 102; DB 1; Length 155;
Best Local Similarity 39.6%; Pred. No. 0.04;
Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

Oy      30 VGAAATGAVGVAGVAGLFGKSGSRVAMAIGGAVLGGLISKSQSDODDKIKL 82
        || | :|||:| | | | | | | | | | | | | | | | | | | | | |
Db       62 VGTLLGGALGIAGTIGCGRGGAIAAIVGAIAGIAGSKIEKMSQVNAEL 114

RESULT 15
YCFU_ECOLI
ID YCFU_ECOLI STANDARD; PRT: 179 AA.
AC P37796; P75951;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
GN Hypothetical protein ycfU.
OS YCFU OR B1110.
OC Escherichia coli.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_Taxid=562;
ON
RN [1] SEQUENCE FROM N.A.
RP MEDLINE-97061202; PubMed=8905232;
RC Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RX Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizochuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Tekemoto K., Wada C., Yamamoto Y.,
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 mln region on the linkage map.";
RL DNABase: 3:137-155(1996).
RM
RN
RP
RX STRAIN-K12;
RC MEDLINE-81236546; PubMed=6265208;
RX Young J.G., Rogers B.L., Campbell H.D., Jaworski A., Shaw D.C.;
RA Nucleotide sequence coding for the respiratory NADH dehydrogenase of
RT Escherichia coli. UUG initiation codon.";
RL Eur. J. Biochem. 116:165-170(1981).
RM [4]
RP IDENTIFICATION
RX MEDLINE-95075659; PubMed=7984428.
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
RL bacterial genome.";
RM Nucleic Acids Res. 22:4756-4767(1994).
CC -/- SIMILARITY: TO RICKETTSIA 17 kDa SURFACE ANTIGEN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
DR EMBL: AE000211; AAC74194.1;
DR EMBL: D00746; BA35525.1;
DR EMBL: V00306; NOT_ANNOTATED_CDS.
```



C:Species: Rickettsia conorii  
C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
C:Accession: B33971  
R:Anderson, B.E.; Tzianabos, T.  
J. Bacteriol. 171, 5199-5201, 1989  
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
A:Reference number: A33971; MUID:89359171  
A:Accession: B33971  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <AND>  
A:Cross-references: GB:M28479; NID:g152463; PIDN:AAA26379.1; PID:g152464  
C:Superfamily: rickettsial common antigen

Query Match  
Best Local Similarity 35.0%; Score 293; DB 2; Length 159;  
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;  
OY 9 SLLIIIV---FLVGC--AONFSRQEVGAATGAVGVAGQLFGKSGRVAMAIGAVLG 63  
D 5 SKIMIALATSMLOACNGPGMKNKOGTGLLGGAGALLGSQFGKQGLVGVGALLG 64  
OY 64 GLISKIGQSMDOODK---IKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYNK 119  
D 65 AVLGGQIGAGMDQDRRLAELTSSQRALETAPSGSNVEMRNPNGNMGVYTPNKTY----- 119  
OY 120 QERRQVCREFQOKAMTAGOKOEITGTACROPDGRMOVIS 159  
D 120 RNSTGYCREYTOTVYVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 3  
A33971  
Rickettsial common antigen precursor - Rickettsia rickettsii  
C:Species: Rickettsia rickettsii  
C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
R:Anderson, B.E.; Tzianabos, T.  
J. Bacteriol. 171, 5199-5201, 1989  
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
A:Reference number: A33971; MUID:89359171  
A:Accession: A33971  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <AND>  
A:Cross-references: GB:M28480; NID:g152457; PIDN:AAA26376.1; PID:g152458  
C:Superfamily: rickettsial common antigen

Query Match  
Best Local Similarity 35.0%; Score 293; DB 2; Length 159;  
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;  
OY 9 SLLIIIV---FLVGC--AONFSRQEVGAATGAVGVAGQLFGKSGRVAMAIGAVLG 63  
D 5 SKIMIALATSMLOACNGPGMKNKOGTGLLGGAGALLGSQFGKQGLVGVGALLG 64  
OY 64 GLISKIGQSMDOODK---IKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYNK 119  
D 65 AVLGGQIGAGMDQDRRLAELTSSQRALETAPSGSNVEMRNPNGNMGVYTPNKTY----- 119  
OY 120 QERRQVCREFQOKAMTAGOKOEITGTACROPDGRMOVIS 159  
D 120 RNSTGYCREYTOTVYVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 4  
G97860  
17k surface antigen precursor [Imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
R:Anderson, B.E.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barde, V.; Samson, D.; Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: G97860  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <KUR>  
A:Cross-references: GB:AEO06914; PIDN:AAL03825.1; PID:g15620425; GSPDB:GN00173  
A:Genes: omp  
C:Superfamily: rickettsial common antigen

Query Match  
Best Local Similarity 35.0%; Score 293; DB 2; Length 159;  
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;  
OY 9 SLLIIIV---FLVGC--AONFSRQEVGAATGAVGVAGQLFGKSGRVAMAIGAVLG 63  
D 5 SKIMIALATSMLOACNGPGMKNKOGTGLLGGAGALLGSQFGKQGLVGVGALLG 64  
OY 64 GLISKIGQSMDOODK---IKLNSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYNK 119  
D 65 AVLGGQIGAGMDQDRRLAELTSSQRALETAPSGSNVEMRNPNGNMGVYTPNKTY----- 119  
OY 120 QERRQVCREFQOKAMTAGOKOEITGTACROPDGRMOVIS 159  
D 120 RNSTGYCREYTOTVYVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 5  
C33971  
Rickettsial common antigen precursor - Rickettsia typhi  
C:Species: Rickettsia typhi  
C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
R:Anderson, B.E.; Tzianabos, T.  
J. Bacteriol. 171, 5199-5201, 1989  
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
A:Reference number: A33971; MUID:89359171  
A:Accession: C33971  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <AND>  
A:Cross-references: GB:M28481; NID:g152459; PIDN:AAA26377.1; PID:g152460  
C:Superfamily: rickettsial common antigen

Query Match  
Best Local Similarity 34.0%; Score 284.5; DB 2; Length 159;  
Matches 55; Conservative 25; Mismatches 49; Indels 9; Gaps 2;  
OY 26 SRQEVGAATGAVGVAGQLFGKSGRVAMAIGAVLGGLISKIGQSMDOODK---IK 81  
D 27 NKQGTGTLGGAGALLGSQFGKQGLVGVGALLGASLDEDRRLLELT 86  
OY 82 LNSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYNKQERRQVCREFQOKAMTAGOKO 141  
D 87 SQRALESAPSGSNVEMRNPNGNMGVYTPNKTY-----RNSTGYCREYTOTVYVIGKQO 141  
OY 142 EIVGTACROPDGRMOVIS 159  
D 142 TTYGNACROPDGMQOVYN 159

RESULT 6  
A25972  
17k antigen precursor - Rickettsia rickettsii  
C:Species: Rickettsia rickettsii  
C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 20-Aug-1999  
R:Anderson, B.E.; Renery, R.L.; Carlone, G.M.; Tzianabos, T.; McNamee, J.E.; Fu, Z.Y.  
J. Bacteriol. 169, 2385-2390, 1987  
A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia rickettsii  
A:Reference number: A25972; MUID:87222152  
A:Accession: A25972

RESULT 10  
S23787  
outer membrane lipoprotein precursor - *Yersinia enterocolitica*  
C:Species: *Yersinia enterocolitica*  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 29-Sep-1999  
C:Accession: S23787  
R:Baeumler, A.J.; Hantke, K.  
J. Bacteriol. 174, 1029-1035, 1992  
A:Title: A lipoprotein of *Yersinia enterocolitica* facilitates ferrioxamine uptake in  
A:Reference number: S23786; MUID:92121089  
A:Accession: S23787  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-155 <BAE>  
 A:Cross-references: EMBL:X50448; NID:g48577; PIDN:CAA4297.1; PID:g48579  
 C:Superfamily: PAL cross-reacting lipoprotein

## Query Match

Best Local Similarity 13.6%; Score 113.5; DB 2; Length 155;  
 Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

QY 10 SLIIISVFLVGCAGN-----FSHQE-----VGAAT 34  
 DB 7 AIAIAATLTGCAANNNTLSGDFSAISDAKOVQTVYTGTLISRPVITGGDDNNVMVGAIG 66  
 QY 35 GAVGVAGVAGQLFGKSGSRVMAAIGAVLIGSKIGOSMDQODKIKL----- 82  
 DB 67 GAVLGGFLGNTVGGGTGSLATATAGAVAGMGAGGAGAMNTGIVOLEVRKDDGTTLIV 126  
 QY 83 --NOSLEKVKAGQVTRWRNDTGSYSVEP 110  
 DB 127 VOKGQPFREFVGO--RVMILASSGVTVVSP 154

## RESULT 11

B82837

Conserved hypothetical protein XF0178 [Imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: B82837  
 R:anonymous  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; PMID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82837  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-257 <SIM>  
 A:Cross-references: GB:AE003872; GB:AE003849; NID:g9104975; PIDN:AAF82991.1; GSPDB:GN001  
 R:Stimpson, A.J.G.; Reimach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 as-Neto, E.; Docena, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carter, H  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzling, J.P.; Krieger, S.E.; Kurano, E.E.; Laigt  
 A:Authors: Martins, E.M.B.N.; Madalira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;  
 A:Authors: da Silva, A.C.R.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 M.; Tshunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0178

## Query Match

Best Local Similarity 13.4%; Score 112; DB 2; Length 257;  
 Matches 20; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 31 GATTGAVGVAGVAGQLFGKSGSRVMAAIGAVLIGLIGSKI 70  
 DB 105 GTACGALITGGLVGNQFGHGNRAKALTRAGAVAGVAGFIGNEV 144

## RESULT 12

AD2696

lipa protein [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: AD2696  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H  
 A:Authors: Gilliet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AD2696  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-142 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AA141986.1; PID:g17739358; GSPDB:GN00186  
 A:Experimental source: Strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: lipa  
 A:Map position: circular chromosome

## Query Match

Best Local Similarity 13.3%; Score 111.5; DB 2; Length 142;  
 Matches 41; Conservative 21; Mismatches 60; Indels 29; Gaps 9;

QY 10 SLIIISVFLVGCAGN-----FSHQE-----VGAAT 34  
 DB 12 SLICVSM-LSAC-----TTGTTPAG--GSLFGR-SAGSTPFLANLGGIGV-K 56  
 QY 70 IQOSMDQODKIKL-----NOSLEKVKAGQVTRWRNDTGSYSVEPRTYORKNOERQ 124  
 DB 57 SCVELDRGDQTKALEVEYKALETAPVGTPIVITGDDVKGGVAMAP---YVGN----- 107  
 QY 125 QYCEPFQOKAMINAGOKOEITGTACROPDGR 155  
 DB 108 QNCRQYSHLTIVDGRDTRVRGAACRNDGSM 138

## RESULT 13

D97478

lipa protein [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
 C:Accession: D97478  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldm  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
 A:Reference number: A87359; PMID:11743194  
 A:Accession: D97478  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-125 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK6781.1; PID:g15155981; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_1782  
 A:Map position: circular chromosome

## Query Match

Best Local Similarity 13.3%; Score 111; DB 2; Length 125;  
 Matches 36; Conservative 18; Mismatches 61; Indels 16; Gaps 6;

QY 30 VAAAGAVGVAGVAGQLFGKSGSRVMAAIGAVLIGLIGSKI 70  
 DB 2 LSACTTGTTRAGOSLFGF-SAPSTPFLANLGGIGV-KSGVELDRGDQTKALEAEYK 59  
 QY 86 LEKVKAGQVTRWRNDTGSYSVEPRTYORKNOERQYCEPFQOKAMINAGOKOEIT 144  
 DB 60 LETAPVGTPIVITGDDVKGGVAMAP---YVGN-----QNCRQYSHLTIVDGRDTRVR 110  
 QY 145 GTACROPDGR 155  
 DB 111 GAACRNDGSM 121

## RESULT 14

S58234

lipa protein - Rhizobium leguminosarum  
 C:Species: Rhizobium leguminosarum



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 ; Search time 8.39378 Seconds

471.414 Million cell updates/sec

Title: US-09-677-374-2

Sequence: 1 MNRGCLGGSSLIISVFLVG.....IYGTACRQPDGRWQVISTEK 162

Scoring table: BLOSUM62

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

```
IssueParents_AA: *
1: /cgn2_6/prodata/1/1aa/5a_COMB.pcp: *
2: /cgn2_6/prodata/1/1aa/5a_COMB.pcp: *
3: /cgn2_6/prodata/1/1aa/6a_COMB.pcp: *
4: /cgn2_6/prodata/1/1aa/6b_COMB.pcp: *
5: /cgn2_6/prodata/1/1aa/PTCUTS_COMB.pcp: *
6: /cgn2_6/prodata/1/1aa/beckfiles1.pcp: *
```

**Pred.** NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81.5	9.7	651	4	US-08-556-978B-19	Sequence 19, App1
2	81.5	9.7	651	4	US-09-247-806-1	Sequence 1, App1
3	81.5	9.7	718	1	US-08-425-069-2	Sequence 2, App1
4	81.5	9.7	718	2	US-08-317-844B-2	Sequence 2, App1
5	81.5	9.7	747	3	US-09-034-177-3	Sequence 3, App1
6	78	9.3	865	1	US-07-803-633A-13	Sequence 13, App1
7	77.5	9.3	255	4	US-09-553-998-8	Sequence 8, App1
8	77	9.2	551	2	US-09-067-351-2	Sequence 2, App1
9	77	9.2	551	4	US-09-360-490-2	Sequence 2, App1
10	76.5	9.2	2516	3	US-08-374-077C-2	Sequence 2, App1
11	76.5	9.2	2516	4	US-08-895-590-2	Sequence 2, App1
12	76	9.1	1021	1	US-07-910-760-12	Sequence 12, App1
13	76	9.1	1021	1	US-08-440-519-12	Sequence 12, App1
14	76	9.1	1021	4	US-08-440-549-12	Sequence 12, App1
15	75	9.0	236	2	US-08-190-199A-65	Sequence 65, App1
16	75	9.0	240	2	US-08-956-047-25	Sequence 25, App1
17	74.5	8.9	913	1	US-08-220-151-6	Sequence 6, App1
18	74.5	8.9	913	1	US-08-413-118-6	Sequence 6, App1
19	74.5	8.9	913	3	US-08-473-446-6	Sequence 6, App1
20	74.5	8.9	913	4	US-09-232-658A-2	Sequence 2, App1
21	74	8.9	970	1	US-08-375-029-7	Sequence 7, App1
22	74	8.9	970	1	US-08-752-929-7	Sequence 7, App1
23	73.5	8.8	970	4	US-09-090-793-5	Sequence 5, App1
24	73.5	8.8	546	2	US-09-067-351-1	Sequence 1, App1
25	73.5	8.8	546	4	US-09-360-490-1	Sequence 1, App1
26	73.5	8.8	733	4	US-08-464-700-2	Sequence 2, App1
27	73	8.7	142	1	US-08-556-823-2	Sequence 2, App1

28	73	8.7	254	4	US-09-128-450-26	Sequence 26, Appl
29	73	8.7	254	4	US-09-0823-494-26	Sequence 26, Appl
30	73	8.7	606	4	US-09-247-805-6	Sequence 6, Appl
31	72.5	8.7	479	4	US-09-177-349-3	Sequence 6, Appl
32	72.5	8.7	547	1	US-08-340-203A-3	Sequence 3, Appl
33	72.5	8.7	547	2	US-08-452-567-3	Sequence 3, Appl
34	72.5	8.7	547	2	US-08-452-427-3	Sequence 3, Appl
35	72.5	8.7	547	3	US-09-085-407-3	Sequence 3, Appl
36	72.5	8.7	731	2	US-08-911-364-1	Sequence 1, Appl
37	72.5	8.7	731	2	US-08-556-978B-10	Sequence 20, Appl
38	72	8.6	101	4	US-08-556-978B-20	Sequence 20, Appl
39	72	8.6	101	4	US-08-556-978B-22	Sequence 22, Appl
40	72	8.6	101	4	US-08-556-978B-32	Sequence 62, Appl
41	72	8.6	101	4	US-09-247-805-5	Sequence 3, Appl
42	72	8.6	101	4	US-09-247-805-5	Sequence 5, Appl
43	72	8.6	101	4	US-09-247-805-7	Sequence 7, Appl
44	72	8.6	604	4	US-08-556-978B-3	Sequence 63, Appl
45	72	8.6	606	4	US-08-556-978B-21	Sequence 21, Appl

## ALIGNMENTS

RESULT 1  
HE-08-556-070B-10

; Sequence 19, Application US/08556978B  
: Patent No. 6269160

GENERAL INFORMATION: ;  
ADDITIONAL COMMENTS: ;

;; TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED  
TITLE OF INVENTION: SPIDER SILK ANALOGS  
.

NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:

ADDRESS: STREET: 1

CITY: WILMINGTON  
STATE: DELAWARE

COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898

COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETT

OPERATING SYSTEM: MICROSOFT WINDOWS

```

; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/5556, 9788B

```

; CLASSIFICATION: 435
;
: PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/077,600  
FILING DATE: JUNE 15, 1993

ATTORNEY/AGENT INFORMATION:  
NAME: FLOYD, LINDA AXAMETH

REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CR-9389-A

TELECOMMUNICATION INFORMATION  
TELEPHONE: 302-892-8112

TELEFAX: 302-773-0164  
INFORMATION FOR SEO ID NO: 19:

SEQUENCE CHARACTERISTICS:  
LENGTH: 651 amino acids

TYPE: amino acid  
STRANDEDNESS: unknown

```

;          TOPOLOGY:  unknown
;
;          MOLECULE TYPE:  protein
;

```

US-08-556-9/88-19

Query match	9.78;
Best Local Similarity	36.88;

malclines 23; conservative 3

28 QEVGA1GAVG---GVAGG---

[illegible]

Db 491 OGAGAAAAVAGGEGIRGAGGCGGCGGGLGCGGAGAAAAAGAGGCG 550  
QY 65 LIGSKIGQ 72  
Db 551 LGGGAGQ 558

RESULT 2  
US-09-247-806-1  
; Sequence 1, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILLIPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/247,806  
; EARLIER FILING DATE: 1999-02-11  
; EARLIER APPLICATION NUMBER: FR 98/01614  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-247-806-1

Query Match  
Best Local Similarity 9.7%; Score 81.5; DB 4; Length 651;  
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 QEVGAATGAVG---GVAGQ-----LFGKSGRVAMATGAVLGG 64  
Db 491 OGAGAAAAVAGGEGIRGAGGCGGCGGGLGCGGAGAAAAAGAGGCG 550  
QY 65 LIGSKIGQ 72  
Db 551 LGGGAGQ 558

RESULT 3  
US-08-425-069-2  
; Sequence 2, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5728810th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/425,069  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-425-069-2

Query Match  
Best Local Similarity 9.7%; Score 81.5; DB 1; Length 718;  
Matches 25; Conservative 3; Mismatches 17; Indels 23; Gaps 2;

QY 28 QEVGAATGAVG---GVAGQ-----LFGKSGRVAMATGAVLGG 64  
Db 491 OGAGAAAAVAGGEGIRGAGGCGGCGGGLGCGGAGAAAAAGAGGCG 550  
QY 65 LIGSKIGQ 72  
Db 551 LGGGAGQ 558

RESULT 4  
US-08-317-844B-2  
; Sequence 2, Application US/08317844B  
; Patent No. 5989894  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5989894th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/317,844B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-105P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-317-844B-2

Query Match 9.7%; Score 81.5; DB 2; Length 718;

```

: APPLICANT: Rudolph, Rainer
: APPLICANT: Schaeffner, Joerg
: APPLICANT: Schwarz, Elisabeth

```

Tue Oct 29 08:30:56 2002

us-09-677-374-2.ra1

Page 4

;; CURRENT APPLICATION NUMBER: US/09/553,498  
;; CURRENT FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: EP9107412.1  
;; PRIOR FILING DATE: 1999-04-26  
;; NUMBER OF SEQ ID NOS: 10  
;; SEQ ID NO: 8  
;; LENGTH: 255  
;; TYPE: PRP  
;; ORGANISM: E. coli  
US-09-553-498-8

Query Match  
Best Local Similarity 20.9%; Score 77.5; DB 4; Length 255;  
Matches 37; Conservative 28; Mismatches 81; Indels 31; Gaps 7;

QY 6 LOGSLIISVFLVGCANFSEQVEGAATGAVV---GGVAGOLFGRKSGRVAMATGAV 61  
DB 83 LQMTSLREEDTAMYYCARDYGAT-MGQGTIVYSSGGSGGGSGGSDIELTSPAI 141  
QY 62 LGLSLGSKIQGSMDDQDKIK-LN-----QSLKVKAGQVTRMRNPDGNSY 106  
DB 142 MSASPGKRVMTCSASSRVYNNMFQOKSGTSPKRWIYDTSKLSGVPARFSGSGSTSY 201  
QY 107 SVEPRTYQRYNKOERROOYCEFOQKMI--AGQKEIYGTACRQPDGRMOVISTE 161  
DB 202 SL-----TISMEADPAATYTCQWSSNPLTFAGTKLEKRAADQ-----KLISEE 249

RESULT 8  
US-09-067-351-2  
; Sequence 2, Application US/09067351  
; Patent No. 5994081  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,351  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-067-351-2

Query Match  
Best Local Similarity 9.2%; Score 77; DB 2; Length 551;  
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 8 GSSLIIISVFLVGCANFSEQVEGAATGAVVAGOLFGRKSGRVAMATGAVLGLIG 67  
DB 54 GASFGSRSLYNLGAKRVSLNGCGSSCRSGFGGRASNGFVNSG---FGYGGVGCGFGSG 110  
QY 68 SK-----IGSMDDQDKIKLNOSLEKVKAGQVTRMRNPDGNSYSEVPRTY 114  
DB 111 PSPFVCPGSDIEVTVNOSLPLRLHIQIDPIQVRAEERQIKTLNKFSTFIDKVFPL 170  
QY 115 QRYNK 119  
DB 171 EQQNK 175

RESULT 9  
US-09-360-490-2  
; Sequence 2, Application US/09360490  
; Patent No. 6221843  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/360,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-360-490-2

Query Match  
Best Local Similarity 9.2%; Score 77; DB 4; Length 551;  
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 8 GSSLIIISVFLVGCANFSEQVEGAATGAVVAGOLFGRKSGRVAMATGAVLGLIG 67  
DB 54 GASFGSRSLYNLGAKRVSLNGCGSSCRSGFGGRASNGFVNSG---FGYGGVGCGFGSG 110

Qy 68 SK-----IGSMDOODKIKLNLNLEKVKAGVTRMRNPDTCNSVSEPVRTY 114  
Db 111 PSPVPCPGIOEVYNQSLITPLHLQIDPTIGRVAREERQIKTLNKTSTIDKVRFL 170  
Qy 115 QRYNK 119  
Db 171 EQQNK 175

RESULT 10  
US-08-374-077C-2  
; Sequence 2, Application US/08374077C  
; Patent No. 6027912  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubald, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha1  
; TITLE OF INVENTION: Calcium Channel Subunit  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374,077C  
; FILING DATE: 19-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm M.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 022650-264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ. ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-374-077C-2

Query Match 9.2%; Score 76.5; DB 3; Length 2516;  
Best Local Similarity 24.8%; Pred. No. 17;  
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VCAATGAVVGVAGOLFEGKSGRYVMAIGG--AVLGGLGSKIGQ---MDQODKIKLN 83  
Db 2389 IGSSNGSIFGSGAGLGGAGSGVG--GLGSSSIRNAFGSGSGPSLSPOHQPYSGTLN 2447  
Qy 84 -----QSELEKVKAGVTRMRNPDTCNSVSEVFRYQRYNKQERRQOYCREFQOKAMLA 137  
Db 2448 SPPIPNRLRLRVATVTYTTNNNKSQVSONSSSLNVRANANSQMNNSPTGQPVQOOSPLR 2507  
Qy 138 GOKOEIYGT 146  
Db 2508 GQGNQTYSS 2516

RESULT 11  
US-08-895-590-2  
; Sequence 2, Application US/08895590  
; Patent No. 6207410

; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubald, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,590  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374,888  
; FILING DATE: 19-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm M.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 022650-263  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ. ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;  
Best Local Similarity 24.8%; Pred. No. 17;  
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VCAATGAVVGVAGOLFEGKSGRYVMAIGG--AVLGGLGSKIGQ---MDQODKIKLN 83  
Db 2389 IGSSNGSIFGSGAGLGGAGSGVG--GLGSSSIRNAFGSGSGPSLSPOHQPYSGTLN 2447  
Qy 84 -----QSELEKVKAGVTRMRNPDTCNSVSEVFRYQRYNKQERRQOYCREFQOKAMLA 137  
Db 2448 SPPIPNRLRLRVATVTYTTNNNKSQVSONSSSLNVRANANSQMNNSPTGQPVQOOSPLR 2507  
Qy 138 GOKOEIYGT 146  
Db 2508 GQGNQTYSS 2516

RESULT 12  
US-07-910-760-12  
; Sequence 12, Application US/07910760  
; Patent No. 5683864  
; GENERAL INFORMATION:  
; APPLICANT: Houghton, Michael  
; APPLICANT: Kuo, Qui-Lim  
; APPLICANT: Kuo, George  
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: P.O. Box 8097 (Int. Prop. R-440)  
; CITY: Emeryville

```

STATE: CA
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,519
FILING DATE: 07-JUL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Blackburn Esq., Robert P.
REGISTRATION NUMBER: 30,447
REFERENCE/DOCKET NUMBER: 0101.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2702
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-910-760-12

Query Match
Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;
Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

QY 31 GAATGAVGVAGAGQLFGK-GSGRVAMAI-----GGAVLGLIGSKI--GQSMDOOKIKLN 83
DB 792 GAATAFVAGAGLAGAIGSVGLKVLIDLAGYAGVAGALVAFKIMSGVPSTEDLVNLL 851
QY 84 QSL-----EKVAKQ-VTRWRN-----PDGNSYSVEPVTRYQRYNKQ 120
DB 852 PALLSPALVGVVCAALIRRHVGPBGAVQVMNRLIAFASRGNHVSPPNSSTNPKPKK 911
QY 121 ERRO-----OYCFEPQCKAMTAGOKOEITYGTAC 148
DB 912 NKRNTNRRODVKFFPGGQIVGVYLLPRGRPLGVKATRTKTSERSOPRGRROP--PKA 969
QY 149 ROPDGR-----WOYISTE 161
DB 970 RRPGRTWAOPGYPMPLYGNE 990

RESULT 13
US-08-440-519-12
Sequence 12, Application US/08440519
Patent No. 5712087
GENERAL INFORMATION:
APPLICANT: Houghton, Michael
APPLICANT: Choo, Qui-Lim
TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)
CITY: Emeryville
STATE: CA
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,519

```

```

FILING DATE: 12-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/910,760
FILING DATE: 07-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blackburn Esq., Robert P.
REGISTRATION NUMBER: 30,447
REFERENCE/DOCKET NUMBER: 0101.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2702
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-519-12

Query Match
Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;
Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

QY 31 GAATGAVGVAGAGQLFGK-GSGRVAMAI-----GGAVLGLIGSKI--GQSMDOOKIKLN 83
DB 792 GAATAFVAGAGLAGAIGSVGLKVLIDLAGYAGVAGALVAFKIMSGVPSTEDLVNLL 851
QY 84 QSL-----EKVAKQ-VTRWRN-----PDGNSYSVEPVTRYQRYNKQ 120
DB 852 PALLSPALVGVVCAALIRRHVGPBGAVQVMNRLIAFASRGNHVSPPNSSTNPKPKK 911
QY 121 ERRO-----OYCFEPQCKAMTAGOKOEITYGTAC 148
DB 912 NKRNTNRRODVKFFPGGQIVGVYLLPRGRPLGVKATRTKTSERSOPRGRROP--PKA 969
QY 149 ROPDGR-----WOYISTE 161
DB 970 RRPGRTWAOPGYPMPLYGNE 990

RESULT 14
US-08-440-549-12
Sequence 12, Application US/08440549
Patent No. 6312889
GENERAL INFORMATION:
APPLICANT: Houghton, Michael
APPLICANT: Choo, Qui-Lim
TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)
CITY: Emeryville
STATE: CA
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,549
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/910,760
FILING DATE: 07-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blackburn Esq., Robert P.

```

```
;
;   REGISTRATION NUMBER: 30,447
;   REFERENCE/DOCKET NUMBER: 0101.002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510) 601-2702
;   TELEFAX: (510) 655-3542
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1021 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-440-549-12

Query Match
Best Local Similarity 20.9%; Pred. No. 6;
Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

OY 31 GAATGAVGVGAGQLFGK-GSGRVAMAI-----GCAVLGLIGSKI--GQSMDOQDKIKLN 83
Db 792 GAATATVAGAGLGAALGAVGLKVLIDILAGYAGVAGALVAFKIMSGEVPSTEDLVNLL 851
OY 84 QSL-----EKVAKQ-VTRRN-----PDTGNSYSVEFPVRYQRYNKQ 120
Db 852 PALISPGALVGVYCAAILRRHVPGEGAVQMMNRILAFASRCNNHVSPCNSSTNPKFOKK 911
OY 121 ERQ-----QYCREFOQKAMIAQKQEIYGTAC 148
Db 912 NKRNTNRPRQDVKPEGGQIVGVYLLPRRGLGVATRKTSERSOPRGRQPI--PKA 969
OY 149 ROPDGR-----MOVISTE 161
Db 970 RREPRTMAQPCYPMPLXNE 990
```

```
RESULT 15
US-08-190-199A-65
; Sequence 65, Application US/08190199A
; Patent No. 5830663
; GENERAL INFORMATION:
;   APPLICANT: EMBLETON, Michael J.
;   APPLICANT: GOROCHOV, Guy
;   APPLICANT: JONES, Peter T.
;   APPLICANT: WINTER, Gregory P.
;   TITLE OF INVENTION: TREATMENT OF CELL POPULATIONS
;   NUMBER OF SEQUENCES: 70
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.
;   STREET: 1100 New York Avenue, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005-3918
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Microsoft Word
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/190,199A
;   FILING DATE: 13-JUL-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/GB92/01483
;   FILING DATE: 10-AUG-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 9212419.7
;   FILING DATE: 11-JUN-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 9117352.6
;   FILING DATE: 10-AUG-1991
;   INFORMATION FOR SEQ ID NO: 65:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 236 amino acids
```

```
;
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-190-199A-65

Query Match
Best Local Similarity 19.6%; Pred. No. 1,1;
Matches 31; Conservative 27; Mismatches 76; Indels 24; Gaps 5;

OY 6 LOGSSLIITISVPLVGAQNFNR---QEVGATGAVGVGAGQLFGKSGSRVAMAI GAVL 62
Db 81 LQMTSLRSEDDTAMYYCARYDGA YWGGTTLVTSAGGGSGGGGGSQIVLTGSPAIM 140
OY 63 GGLIGSKIGSQMDQDKIK-LN-----QSLKVRAGQVTRRNPDGNSYS 107
Db 141 SASPGEKVTWTCASSSVRYMNFQOKSGTSPKRWIYDTSKSSGVAPRPSGSGSTYS 200
OY 108 VEPVRYQRYNKOERRQOYCREFOQKAMI--ACOKOEI 143
Db 201 L----TISMEADADATYYCQWSSNPLTFGAGTKLEL 234
```

Search completed: October 27, 2002, 11:00:27  
Job time : 10.3938 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 21.8238 Seconds  
(without alignments)  
824,309 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836  
Sequence: 1 MNRGCLGSGSLIIISVPLVG.....IYGFACRQPDGRWQVISTEK 162

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	836	100.0	162	22	AA678025	Piscirickettsia sa
2	836	100.0	162	22	AA681126	Ospa antigen amino
3	815	97.5	161	22	AA681127	Optimised Ospa pro
4	815	97.5	256	22	AA681128	CITE2 Ospa Constru
5	112	13.4	20	22	AA681130	Ospa B-cell epitop
6	108	12.9	224	22	AA620105	Moraxella catarrha
7	105.5	12.6	223	20	AA34487	Porphyromonas ging
8	105.5	12.6	220	20	AA34362	Porphyromonas ging
9	102	12.2	154	11	AA605799	PBOMP-2 gene prod.
10	98	11.7	309	22	AA615906	Novel human diagno
11	91	10.9	528	22	AA682611	Spider recombinant

12	86.5	10.3	2309	22	ABB66232	Drosophila melanog
13	82	9.8	666	22	ABB58019	Drosophila melanog
14	81.5	9.7	651	20	AAV40097	Spider silk proteol
15	81.5	9.7	718	12	AA614308	N. clavipes draglin
16	81.5	9.7	718	19	AA653346	Nephila clavipes s
17	81.5	9.7	718	21	AAV59070	N. clavipes spider
18	80.5	9.6	102	22	AA641943	Human polyptide
19	80.5	9.6	302	22	AA640157	Human polyptide
20	80.5	9.6	542	22	ABB65790	Drosophila melanog
21	80.5	9.6	542	22	ABB65791	Drosophila melanog
22	80.5	9.6	542	22	ABB70501	Drosophila melanog
23	80	9.6	116	19	AAV11028	H. pylori ORF 01cp
24	80	9.6	2017	22	ABG06301	Novel human diagno
25	80	9.6	2599	21	AAV75098	Neisseria meningit
26	79	9.4	618	21	AA656803	Human prostate can
27	78	9.3	646	18	AA627178	Nephila clavipes s
28	78	9.3	865	14	AA630169	Marek's Disease VI
29	77.5	9.3	255	21	AA611398	E. coli expression
30	77.5	9.3	255	22	AA674199	Pe1B-scFvOxazoln
31	77.5	9.3	255	22	AA670769	Expression plasmid
32	77.5	9.3	255	22	AAV72020	E. carotovora Pe1B
33	77.5	9.3	900	22	ABG28648	Novel human diagno
34	77	9.2	147	22	ABB69847	Drosophila melanog
35	77	9.2	429	10	AA690064	Antigenic protein
36	77	9.2	551	21	AAV52398	Human keratin KERT
37	77	9.2	1251	21	ABB61254	Drosophila melanog
38	76.5	9.2	514	22	AA636520	Pseudomonas aerugi
39	76.5	9.2	2516	17	AA601875	Neutromal Invertebr
40	76.5	9.2	2516	17	AA601884	Invertebrate calci
41	76.5	9.2	2516	22	ABB61076	Drosophila melanog
42	76	9.1	300	11	AA605878	Mezozile surface a
43	76	9.1	324	22	AA676751	Corynebacterium gl
44	76	9.1	346	22	AA692057	C glutamicum prote
45	76	9.1	346	22	AA676750	Corynebacterium gl

#### ALIGNMENTS

RESULT 1	
ID	AA678025 standard; Protein; 162 AA.
XX	
AC	AA678025.
XX	
DT	15-JAN-2002 (first entry)
XX	
DE	Piscirickettsia salmonis polypeptide p10.6.
XX	
KW	Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW	septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW	ATCC VR-1361.
XX	
OS	Piscirickettsia salmonis.
XX	
PN	W0200168865-A2.
XX	
PD	20-SEP-2001.
XX	
PF	12-MAR-2001; 2001WO-GB01055.
XX	
PR	11-MAR-2000; 2000GB-0005838.
PR	01-JUL-2000; 2000GB-0016080.
PR	01-JUL-2000; 2000GB-0016082.
PR	29-JUL-2000; 2000GB-0018599.
PA	(AQUA-) AQUA HEALTH EURO LTD.
XX	
PI	Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI	Burzio L;
XX	
DR	WPI: 2001-639050/73.
DR	N-PDB; AA679040.

XX New nucleic acids encoding an amino acid sequence homologous to the  
PT surface antigen present on piscirickettsia salmonis are useful to  
PT protect fish against piscirickettsiosis

PS Claim 6; Fig 5; 25pp; English.

CC The invention relates to nucleic acid sequences and the encoded protein  
CC of a least part of the surface antigen present on piscirickettsia  
CC salmonis for production of a vaccine with antibacterial activity to  
CC protect fish against P. salmonis which causes piscirickettsiosis, also  
CC known as salmonid rickettsial septicaemia.

XX Sequence 162 AA:

Query Match 100.0%; Score 836; DB 22; Length 162;  
Best Local Similarity 100.0%; Pred. No. 3.2e-80;  
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLQGSSLLITISVFLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMATGCA 60  
DB 1 MNRGCLQGSSLLITISVFLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMATGCA 60  
QY 61 VLGLIGSKIGQSMDOQDKIKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120  
DB 61 VLGLIGSKIGQSMDOQDKIKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120  
QY 121 ERROQYCREFOQKAMTAGOKOEITGTACROPDGRMQVISTEK 162  
DB 121 ERROQYCREFOQKAMTAGOKOEITGTACROPDGRMQVISTEK 162

RESULT 2

ID AAB81126 standard; Protein; 162 AA.  
XX AAB81126;

AC AAB81126;  
XX 11-JUL-2001 (first entry)

DE Ospa antigen amino acid sequence.

XX Polkillothermic fish; piscirickettsia salmonis; rickettsial pathogen;  
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
XX SRS.

XX Piscirickettsia salmonis.

OS Piscirickettsia salmonis.  
FH Key Location/Qualifiers  
FT Region 110..129  
XX /label=B\_cell\_epitope

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

XX (KAYW/) KAY W W.  
PA (BURI/) BURIAN J.  
XX (KUZV/) KUZV M A.

PI Kay WW, Burian J, Kuzv MA;

DR WPI; 2001-316844/34.

DR N-PSDB; AAF86246.

PT Method for protecting polkillothermic fish against salmonid rickettsial  
PT septicaemia and other rickettsial diseases comprises administering a  
XX vaccine containing the Ospa protein of piscirickettsia salmonis

PS Example 2; Fig 2B; 35pp; English.

CC This invention relates to a method for the protection against infection  
CC of a polkillothermic fish by the bacterial pathogen, piscirickettsia  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a  
CC Ospa in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkillothermic fish, against the bacterial pathogen  
CC P. salmonis. The method is also useful for protecting against salmonid  
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
CC sequence represents P. salmonis Ospa protein. An Ospa protein with an  
CC antibody response.

XX Sequence 162 AA:

Query Match 100.0%; Score 836; DB 22; Length 162;  
Best Local Similarity 100.0%; Pred. No. 3.2e-80;  
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLQGSSLLITISVFLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMATGCA 60  
DB 1 MNRGCLQGSSLLITISVFLVCAQNFSSROEVGAATGAVGVAGOLFSGSGRVAMATGCA 60  
QY 61 VLGLIGSKIGQSMDOQDKIKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120  
DB 61 VLGLIGSKIGQSMDOQDKIKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQ 120  
QY 121 ERROQYCREFOQKAMTAGOKOEITGTACROPDGRMQVISTEK 162  
DB 121 ERROQYCREFOQKAMTAGOKOEITGTACROPDGRMQVISTEK 162

RESULT 3

ID AAB81127 standard; Protein; 161 AA.  
XX AAB81127;

AC AAB81127;  
XX 11-JUL-2001 (first entry)

DE Optimised Ospa protein 17E2 amino acid sequence.

XX Polkillothermic fish; piscirickettsia salmonis; rickettsial pathogen;  
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
XX SRS.

XX Piscirickettsia salmonis.

OS Synthetic.

FH Key Location/Qualifiers  
FT Region 109..128  
XX /label=B\_cell\_epitope

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

XX (KAYW/) KAY W W.  
PA (BURI/) BURIAN J.  
XX (KUZV/) KUZV M A.

PI Kay WW, Burian J, Kuzv MA;

DR WPI; 2001-316844/34.

DR N-PSDB; AAF86247.

PT Method for protecting polkillothermic fish against salmonid rickettsial  
PT septicaemia and other rickettsial diseases comprises administering a

PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -  
XX  
PS Disclosure: Fig 5; 35pp; English.  
CC  
CC This invention relates to a method for the protection against infection  
CC of a polkillothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC *salmonis*. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of a  
CC Ospa in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkillothermic fish, against the bacterial pathogen  
CC *P. salmonis*. The method is also useful for protecting against salmonid  
CC rickettsial septicæmia (SRS) and other rickettsial diseases. The present  
CC sequence represents optimised *P. salmonis* Ospa protein 17E2. The DNA  
CC encoding Ospa 17E2 (AA86247) has been optimised for expression in  
CC *Escherichia coli*. An Ospa protein with an N-terminal fusion partner is  
CC used in a vaccine to create an anti-Ospa antibody response.  
XX  
SQ Sequence 161 AA:  
Query Match 97.5%; Score 815; DB 22; Length 161;  
Best Local Similarity 98.8%; Pred. No. 5,1e-78;  
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 3 RGLGSSLLIITISVFLVGCANFSROEVGAATGAVVGVAGOLFSGKSGRVANAIGAVL 62  
DB 2 RGLGSSLLIITISVFLVGCANFSROEVGAATGAVVGVAGOLFSGKSGRVANAIGAVL 61  
OY 63 GGLIGSKIGSMQDQDKIKLNOSLEKVKAGVTRWRNPDTGNSYSVEPRTYORYNKQER 122  
DB 62 GGLIGSKIGSMQDQDKIKLNOSLEKVKAGVTRWRNPDTGNSYSVEPRTYORYNKQER 121  
OY 123 ROOYCREFOOKAMTAGOKETIGTACROPDGRMOVISTEK 162  
DB 122 ROOYCREFOOKAMTAGOKETIGTACROPDGRMOVISTEK 161  
RESULT 4  
AAB81128  
ID AAB81128 standard; Protein; 256 AA.  
XX  
AC AAB81128;  
XX  
DT 11-JUL-2001 (first entry)  
XX  
DE C17E2 Ospa construct with N-terminal fusion partner.  
XX  
KW Polkillothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;  
KW vaccine; Ospa; salmonid rickettsial septicæmia; rickettsial disease;  
KW SRS; 17E2; fusion construct.  
XX  
OS *Piscirickettsia salmonis*.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 1..95  
FT /label= Undefined\_N-terminal\_fusion\_partner  
FT 96..256  
FT /label= C17E2\_Ospa  
FT /note= "Product of Ospa gene optimised for expression in  
FT *Escherichia coli*".  
XX  
XX CA2281913-A1.  
XX  
XX 17-MAR-2001.  
XX  
XX 17-SEP-1999; 99CA-2281913.  
XX  
XX 17-SEP-1999; 99CA-2281913.  
XX  
XX 17-SEP-1999; 99CA-2281913.  
XX  
XX (KAYW/) KAY W W.  
XX (BURI/) BURIAN J.  
XX (KUZV/) KUZYK M A.  
XX

PI Kay WM, Burian J, Kuzyk MA;  
XX  
DR WPI: 2001-316844/34.  
XX  
DR N-PSDB; AAF86248.  
XX  
PT Method for protecting polkillothermic fish against salmonid rickettsial  
PT septicæmia and other rickettsial diseases comprises administering a  
PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -  
XX  
XX Example 4; Fig 5; 35pp; English.  
XX  
PS  
XX This invention relates to a method for the protection against infection  
XX of a polkillothermic fish by the bacterial pathogen, *Piscirickettsia*  
XX *salmonis*. The method comprises administering an immunogenic amount of a  
XX *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of a  
XX Ospa in the form of a vaccine. The method is used for protecting animals,  
XX particularly polkillothermic fish, against the bacterial pathogen  
XX *P. salmonis*. The method is also useful for protecting against salmonid  
XX rickettsial septicæmia (SRS) and other rickettsial diseases. The present  
XX sequence represents the amino acid sequence of C17E2, a *P. salmonis* Ospa  
XX construct optimised for expression in *Escherichia coli*, fused to an  
XX undefined N-terminal fusion partner. The fusion protein is used in a  
XX vaccine to create an anti-Ospa antibody response.  
XX  
SQ Sequence 256 AA:  
Query Match 97.5%; Score 815; DB 22; Length 256;  
Best Local Similarity 98.8%; Pred. No. 9,3e-78;  
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 3 RGLGSSLLIITISVFLVGCANFSROEVGAATGAVVGVAGOLFSGKSGRVANAIGAVL 62  
DB 97 RGLGSSLLIITISVFLVGCANFSROEVGAATGAVVGVAGOLFSGKSGRVANAIGAVL 156  
OY 63 GGLIGSKIGSMQDQDKIKLNOSLEKVKAGVTRWRNPDTGNSYSVEPRTYORYNKQER 122  
DB 157 GGLIGSKIGSMQDQDKIKLNOSLEKVKAGVTRWRNPDTGNSYSVEPRTYORYNKQER 216  
OY 123 ROOYCREFOOKAMTAGOKETIGTACROPDGRMOVISTEK 162  
DB 217 ROOYCREFOOKAMTAGOKETIGTACROPDGRMOVISTEK 256  
RESULT 5  
AAB81130  
ID AAB81130 standard; Peptide; 20 AA.  
XX  
AC AAB81130;  
XX  
DT 11-JUL-2001 (first entry)  
XX  
DE Ospa B-cell epitope peptide #2.  
XX  
KW Polkillothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;  
KW vaccine; Ospa; salmonid rickettsial septicæmia; rickettsial disease;  
KW SRS; antibody.  
XX  
OS *Piscirickettsia salmonis*.  
OS  
XX  
XX CA2281913-A1.  
XX  
XX 17-MAR-2001.  
XX  
XX 17-SEP-1999; 99CA-2281913.  
XX  
XX 17-SEP-1999; 99CA-2281913.  
XX  
XX 17-SEP-1999; 99CA-2281913.  
XX  
XX (KAYW/) KAY W W.  
XX (BURI/) BURIAN J.  
XX (KUZV/) KUZYK M A.  
XX  
XX Kay WM, Burian J, Kuzyk MA;  
XX

DR WPI: 2001-316844/34.

XX Method for protecting poikilothermic fish against salmonid rickettsial  
PT septicemia and other rickettsial diseases comprises administering a  
XX vaccine containing the Ospa protein of *Piscirickettsia salmonis* -  
PS Example 2; Page 17; 35pp; English.

XX This invention relates to a method for the protection against infection  
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic amount of a  
CC Ospa in the form of a vaccine. The method is used for protecting fragment of a  
CC particularly poikilothermic fish, against the bacterial pathogen  
CC rickettsial septicemia (SR) and other rickettsial diseases. The present  
CC sequence represents an immunogenic epitope of the *P. salmonis* Ospa  
CC protein. The peptide is used to raise rabbit anti-Ospa antibodies.  
SQ Sequence 20 AA;

Query Match 13.4%; Score 112; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.9e-05;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 110 PVRTGQRYKQRRQGYCRE 129  
Db 1 PVRTGQRYKQRRQGYCRE 20  
|||||

## RESULT 6

ID AAB20105 standard; Protein; 224 AA.  
XX AAB20105;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB113 protein.

XX BASB113: infection; otitis media; pneumonia; therapy; diagnosis;  
KW antibacterial; antimicrobial.

OS Moraxella catarrhalis.

PN WO200100836-A1.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-EP05851.

PR 25-JUN-1999; 99GB-0015044.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonnard J;

XX WPI: 2001-112458/12.

DR N-PSDB; AAF30043.

XX New BASB113 polypeptide isolated from Moraxella catarrhalis bacterium,  
PT useful for diagnosing and producing vaccines against bacterial  
PT infections such as otitis media and pneumonia  
PS Claim 1; Page 67; 86pp; English.

XX The present sequence is that of BASB113 protein from Moraxella  
CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis  
CC media in children and pneumonia in adults. The invention provides  
CC BASB113 polypeptides, and polynucleotides encoding them, as well as  
CC expression vectors, host cells and methods for producing BASB113  
CC polypeptides using recombinant methods. Also claimed is a vaccine  
CC composition comprising a BASB113 polypeptide, an immunogenic

CC fragment of a BASB113 polypeptide, or a polypeptide having at least  
CC 85% amino acid sequence identity to BASB113, or comprising a  
CC polynucleotide encoding such a polypeptide. A claimed method of  
CC diagnosing a Moraxella infection involves identifying a BASB113  
CC polypeptide or antibody. A claimed therapeutic composition useful  
CC in treating humans with M. catarrhalis infection comprises at least  
CC 1 antibody directed against a BASB113 polypeptide. BASB113  
CC polypeptides also have utility in raising specific antibodies.  
XX and in screening for antibacterial drugs.

SQ Sequence 224 AA;

Query Match 12.9%; Score 108; DB 22; Length 224;  
Best Local Similarity 32.4%; Pred. No. 0.0028;  
Matches 36; Conservative 16; Mismatches 35; Indels 24; Gaps 5;

OY 8 GSSLITIVFLVGCANFRQEVGATG-----AVGCVAGQLFGKSGRVAMATG 59  
Db 7 GVLLASSMALAGCANVTG---TGNGTFGCVANVKAIVGAVAGAL---GGRATSKATGG 60  
OY 60 -----AVLGGIGSKIGSQSDQDKIKLNQSLKVKAGQVTRMRNPTGN 104  
Db 61 EKTGRDAITLGAAGAAGAVMERQAK-----QIEQWQGTGYVTHIDTGN 107  
|||||

## RESULT 7

ID AAY34487 standard; Protein; 223 AA.  
XX AAY34487;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG3.

XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
KW vaccine; antigenic.

OS Porphyromonas gingivalis.

PN WO9929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSL-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
PI Ross BC, Rothel LJ, Webb EA;

DR WPI: 1999-385613/32.

DR N-PSDB; AAX91705.

XX Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis  
XX

PS Claim 1; Page 469; 588pp; English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the



DB

62 VCTLGGGALGIGSTIGGGRGQALAAVGAIGALGAKSIEKMSQVNGAEI 114

RESULT 10

ABG15906

ID ABG15906 standard; Protein; 309 AA.

AC ABG15906;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15897.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN MO200175067-A2.

PD 11-OCT-2001.

PE 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS80093.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT biodiversity for genetic disorders or other traits and to assess

PS Claim 20; SEQ ID No 46265; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques (II). (II) is useful for generating antibodies against it, detecting or a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations in responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences. ABG00010-ABG30377 represent novel human Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pot\_sequences.

XX Sequence 309 AA;

Query Match

Best Local Similarity 11.7%; Score 98; DB 22; Length 309;

Matches 33; Conservative 11; Mismatches 39; Indels 14; Gaps 3;

QY 13 IISVFLVGCANQNSR-----QEVGATGAVVGVAGOLFQK--SGRVMAAIGAVLGL 65

pb 185 IISILPAKVAVNSQNRKRAQAFGALGAVAGVIGHNVSGSNGITTAAGVGAAGVAA 244  
 J

QY 66 TSGKI-----GQMDQODKIKLNLSLEKVRKAGQVT 95  
 Db 245 AGSMVNDKTLMEGVSLTYRKGTQVTSYOVGKEEFT 281

RESULT 11

AAB82611

ID AAB82611 standard; Protein; 528 AA.

AC AAB82611;

DT 02-OCT-2001 (first entry)

DE Spider recombinant silk protein PETNCDs.

KW Spider; orb-weaver; silk protein; PETNCDs; structural protein;

KM purification; fibre; spinning.

OS Nephila clavipes.

FH Key Location/Qualifiers

FT Misc-difference 417 /note= "encoded by TAA"

FT Misc-difference 427 /note= "encoded by CGA"

FT Misc-difference 522 /note= "encoded by GAG"

PN MO200153333-A1.

PD 26-JUL-2001.

PE 01-NOV-2000; 2000WO-US30086.

PR 20-JAN-2000; 2000US-0490291.

XX (MEL/) MELLO C M.

PA (ARCI/) ARCIDIACONO S.

PA (BUTL/) BUTLER M M.

PA (USAA ) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

DR WPI; 2001-483136/52.

DR N-PSDB; AAH26304.

Recovering structural polypeptides in a biological sample, useful for purifying and spinning spider silks and other structural proteins, comprises treating the sample containing the polypeptides with an acid

PS Claim 2; Page 41-42; 49pp; English.

The present sequence is that of the orb-weaver spider (Nephila clavipes) recombinant silk protein PETNCDs. The invention provides methods for purifying and spinning spider silks and other structural proteins. Organic acids are used to lyse recombinant cells or other biological samples (such as non-recombinantly derived cells), and enrich the purity and yields of structural proteins by hydrolysing many of the macromolecules while leaving the structural proteins intact. In the case of silk proteins, the resulting lysate is further purified by ion-exchange or affinity chromatography and processed into an aqueous-based mixture for into vector pET24 for recombinant expression in *Escherichia coli*, containing denaturant (3 M guanidine-HCl) and by affinity chromatography on nickel-NTA agarose. Products obtained using the methods of the invention can be used in the construction of many ballistic protection, parachutes and parachute cords. The new method has the following advantages over prior art: it involves fewer steps, requires less time and smaller volumes of reagents,

CC results in better recovery of protein at higher purity (70-99%),  
CC is easy to scale up, and the fibres are spun in an environmentally  
CC benign solution reducing hazardous waste accumulation and cost.

XX Sequence 528 AA:

Query Match 10.9%; Score 91; DB 22; Length 528;  
Best Local Similarity 51.1%; Pred. No. 0.51;  
Matches 24; Conservative 2; Mismatches 19; Indels 2; Gaps 1;

Oy 28 QEVGAATGAVGVACQ--LFGKSGRVAITAIGAVLGLISKIQ 72  
DB 320 QGAGAAAAAAGAGCGGCGAGAAAAAAGAGCGGCGAGQ 366

RESULT 12  
ABB66232

ID ABB66232 standard; Protein; 2309 AA.

XX ABB66232;

AC 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 25488.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI: 2001-656860/75.  
N-PSDB: ABL10335.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Disclosure: SEQ ID NO 25488; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2309 AA:

Query Match 10.3%; Score 86.5; DB 22; Length 2309;  
Best Local Similarity 27.4%; Pred. No. 9.9;  
Matches 34; Conservative 18; Mismatches 55; Indels 17; Gaps 5;

Oy 23 QNFROEVAATGA---VVGAVAGOLFGRGSGRVAMAIGAVLGLISKIGOSMDQDK 79  
DB 2124 QQQQQRQVGGGNPSNMALJGREGSGAVSGSGN-----GGGGGCGAGSGVGGNGGNGCY 2178

Oy 80 IKLNSLEKVKAGQVTRMNP-DTGNYSVEPYRYQRYNKOERRQYCRFQKAMING 138  
DB 2179 GSVGDS---GGGGGCRPTPTIQRPNNTPOHP-----Q00000000000REDAAAAAVA 2230

Oy 139 OKOE 142  
DB 2231 QRAG 2234

RESULT 13  
ABB58019

ID ABB58019 standard; Protein; 666 AA.

XX ABB58019;

AC 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 849.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI: 2001-656860/75.  
N-PSDB: ABL02122.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Disclosure: SEQ ID NO 849; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 666 AA:

Query Match 9.8%; Score 82; DB 22; Length 666;  
Best Local Similarity 22.1%; Pred. No. 6.1;  
Matches 30; Conservative 22; Mismatches 58; Indels 26; Gaps 2;

Oy 31 GATGAVGVAGOLFGRGSGRVAMAIGAVLGLISKIGOSMDQDKIKLNSLEKVK 90  
DB 463 GQPTGAGGGGGSSGKNAGTILAVS-GSVTTGAGAGSGTGSSNSNSGSGSVSGSV 521  
Oy 91 AGQVTRMNPDTGNSYSV-----EPVRYQRYNKOERRQ 125  
DB 522 SGSGSHPTPTSLHAHSANGTSSLLGGDLATPHQVVAAGSPVWLHQ000000000 581





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:01 ; Search time 18.1865 Seconds

(without alignments)  
1540.986 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836  
Sequence: 1 MNRGCLQGSSLIITISVFLVG.....ITYGTRQRPDGRWQVISTEK 162

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_ricent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirts:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	836	100.0	162	2	09F9K8
2	313.5	37.5	148	2	054381
3	291	34.8	159	2	09F9F2
4	266.5	31.9	137	2	052252
5	260.5	31.2	144	2	09K2N6
6	259.5	31.0	137	2	031065
7	259.5	31.0	144	2	09K4W8
8	252	30.1	154	2	053154
9	247	29.5	151	2	09F909
10	237.5	28.4	131	2	09F001
11	237.5	28.4	131	2	052637
12	236.5	28.3	131	2	09L522
13	223.5	26.7	105	2	031208
14	143	17.1	199	16	0985G4
15	140	16.7	77	2	09AGC7
16	130.5	15.6	136	16	092R89

17	128.5	15.4	182	16	09HXT3	09hxt3 pseudomonas
18	116	13.9	154	16	09I4S1	09i4s1 pseudomonas
19	112	13.4	257	16	09PGX0	09pgx0 xyliella fas
20	111	13.3	139	2	052854	052854 rhizobium 1
21	110.5	13.2	232	16	09A3X8	09a3x8 caulobacter
22	108.5	13.0	155	2	09F6B1	09f6b1 edwardsiell
23	107.5	12.9	153	2	069776	069776 rhizobium e
24	107.5	12.9	155	2	09RA95	09ra95 serratia sp
25	106.5	12.7	221	16	092ST9	092st9 rhizobium m
26	106	12.7	155	2	09RB08	09rb08 pectobacter
27	105.5	12.6	223	2	09XC44	09xc44 porphyromon
28	101.5	12.1	304	16	09I762	09i762 pseudomonas
29	99.5	11.9	83	16	092LP2	092lp2 rhizobium m
30	99	11.8	154	16	09CNR3	09cnr3 pasteurilla
31	98	11.7	172	16	P76572	P76572 escherichia
32	98	11.7	220	2	09YAX8	09yax8 pseudomonas
33	96	11.5	223	16	09KSR1	09ksr1 vibrio chol
34	95.5	11.4	79	16	092NT4	092nt4 rhizobium m
35	94.5	11.3	838	2	09AL49	09al49 shigella fl
36	94	11.2	608	10	09SUX1	09sux1 arabidopsis
37	93.5	11.2	257	16	09A8M8	09a8m8 caulobacter
38	92	11.0	105	16	098P93	098p93 rhizobium 1
39	91	10.9	544	5	046171	046171 nephila cla
40	90	10.8	105	16	0983Y0	0983y0 rhizobium 1
41	89	10.6	137	16	09HU07	09hu07 pseudomonas
42	88.5	10.6	161	2	068681	068681 bacillus me
43	88.5	10.6	216	16	098NA9	098na9 rhizobium 1
44	87.5	10.5	407	16	031557	031557 bacillus su
45	87	10.4	201	16	0930X4	0930x4 rhizobium m

#### ALIGNMENTS

RESULT 1  
ID 09F9K8 PRELIMINARY; PRT: 162 AA.  
AC 09F9K8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE 17 KDA ANTIGEN.  
GN OSPA.  
OS Piscirickettsia salmonis.  
OC Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;  
OC Piscirickettsia.  
OX NCBI\_TaxID=1238;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LF-89;  
RA Kuzyk M.A., Burian J., Thornton J.C., Kay W.W.;  
RT "Identification of a genus-common Rickettsial surface antigen in the  
RT salmonid pathogen Piscirickettsia salmonis.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF184152; ANGI7000.1; -  
SQ SEQUENCE 162 AA; 17661 MW; DDE39BEFD94A527E CRC64;

Query Match 100.0%; Score 836; DB 2; Length 162;

Best local similarity 100.0%; Pred. No. 9.6e-68;

Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGCLQGSSLIITISVFLVGCAONFSPROEYGAATGAVGCVAGOLFGKSGRYAAMIGGA 60  
DB 1 MNRGCLQGSSLIITISVFLVGCAONFSPROEYGAATGAVGCVAGOLFGKSGRYAAMIGGA 60  
QY 61 VLGGILGSKTIGQSMDOODKTKLNQSLSEKYGAGOVTRMRNPDTGNSYSVEPRTYORYNKO 120  
DB 61 VLGGILGSKTIGQSMDOODKTKLNQSLSEKYGAGOVTRMRNPDTGNSYSVEPRTYORYNKO 120  
QY 121 ERRQOYCRFQOKRAMIAGOKOEIYGTACROPDGRWQVISTEK 162  
DB 121 ERRQOYCRFQOKRAMIAGOKOEIYGTACROPDGRWQVISTEK 162

```

OY 64 GLIGSKIGQSMDOOK----IKINOSLEKYGAGVTRMRNPDGNSYSVEPVRTYORYNK 119
Db 65 AVLLGGIGAGNDEEDDRRLAELTSGRALERTPGSTVEWRNPDNGHGYTPNKTY----- 119
OY 120 QERRQOYCRFEQKAMIAAGKDEITGTACROPDRHWQVYS 159
Db 120 RNSTGYCREYTOTVYIGGKQOKAYGNACROPDGIHWQVYN 159

RESULT 4
052252
ID 052252 PRELIMINARY; PRT; 137 AA.
AC 052252;
DT 01-JUN-1998 (TREMBLrel, 06, Created)
DT 01-JUN-1998 (TREMBLrel, 06, Last sequence update)
DE 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
OS 17 KDA ANTIGEN (FRAGMENT).
OS Rickettsia cooley.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxId=69410;
RN 1
RP SEQUENCE FROM N.A.
RA Billings A.N., Teltow G.J., Walker D.H.;
RT "Molecular characterization of a novel spotted fever group rickettsial
RT species from Ixodes scapularis in Texas."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF031534; AAB95267.1; -.
FT NON_TER 1
SQ SEQUENCE 137 AA; 14215 MW; A27597A9AFD85FC3 CRC64;

Query Match
Best Local Similarity 31.9%; Score 266.5; DB 2; Length 137;
Matches 53; Conservative 23; Mismatches 49; Indels 9; Gaps 2

OY 23 QNFSRQEVGATGAVVGVGAQGFQKSGSRVMAIGAVLGLGSKIGQSMDOOK--- 79
Db 7 RGNMKQGTGTLLGAGCALLGSGQFGKGGQLGVGVGALLGAVLGIGAGNDEEDDRRLA 66
OY 80 -IKINOSLEKYGAGVTRMRNPDGNSYSVEPVRTYORYNKQERRQOYCRFEQKAMIA 138
Db 67 ELISQALNAASGSGSTWRNPDNGNMGVTPNKTY-----RNSTGYCREYTOTVYIG 121
OY 139 OKREITGYACROPD 152
Db 122 KQOKAYGNACROPD 135

RESULT 5
09K2N6
ID 09K2N6 PRELIMINARY; PRT; 144 AA.
AC 09K2N6;
DT 01-OCT-2000 (TREMBLrel, 15, Created)
DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)
DE 17KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT)
OS male-killing Rickettsia from Adalia bipunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxId=38028;
RN 1
RP SEQUENCE FROM N.A.
RA Schlenker H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.;
RT "On the evolution of male-killing Rickettsia (a-Proteobacteria) from two con-
RT generic ladybirds, Adalia bipunctata L. and A. decempunctata L.
RT (Coleoptera: Coccinellidae)."
DL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ269518; CAB96383.1; -.
FT EMBL; AJ269517; CAB96382.1; -.
FT NON_TER 1
SQ SEQUENCE 144 AA; 14415 MW; A27597A9AFD85FC3 CRC64;

```



Db 125 QD-----CRYTGTVTGKQKAKYGNACRQPDGQ 154

## RESULT 9

Q9F909

AC Q9F909 PRELIMINARY; PRT; 151 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE OUTER MEMBRANE PROTEIN (FRAGMENT)

OS Rickettsia helvetica

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia

NCBI\_TaxID=35789;

RP SEQUENCE FROM N.A.

RA Nilsson K., Paulson C.;

RT "Novel peptide diagnostic reagent and kit for detection of

RL rickettsiosis."

DR Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.

FT EMBL: AF181036; AAG09427.1; -

NON\_TER 151

SEQUENCE 151 AA: 15621 MW; B77407B9C71E4B39 CRC64;

Query Match

Best Local Similarity 29.5%; Score 247; DB 2; Length 151;

Matches 53; Conservative 26; Mismatches 57; Indels 14; Gaps 4;

QY 9 SLLIIV---FLVGC--AQNRSREVGATGAVGAGVAGQLFGKSGRVAMAIGAVIG 63

Db 5 SKIMITALAASMLQACNGPGKNGKOGTGTLLGAGAGALLSGQKGGQGLVGVGALLG 64

QY 64 GLIGSKIQSDQDK---IKLNSLEKVVAGQVTRWNPDTGNSYSVEPVRTYQRYNK 119

Db 65 AVLGQIYAGDEDDRLAELTSGRALEAPSGSNVEMRPDNGYVTPNKTY----- 119

QY 120 QERRQCYREFQOKAMTAGOQKQELGYACROP 151

Db 120 RNSTGYCREYQTVVIGGKQKAKYGNACROP 151

RESULT 10

Q9F001

AC Q9F001 PRELIMINARY; PRT; 131 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 17 KDA PROTEIN (FRAGMENT)

OS Rickettsia sp. California 2.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia

NCBI\_TaxID=147259;

RP SEQUENCE FROM N.A.

RA Raoult D.;

RT "A new Sfg rickettsia isolated from fleas."

RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Roux V., Raoult D.;

RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

FT EMBL: AF210693; AAG48554.1; -

NON\_TER 131

SEQUENCE 131 AA: 13374 MW; 23C8819B29FEFF60 CRC64;

Query Match

Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;

Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGAGVAGQLFGKSGRVAMAIGAVLGLIGSKIGQSDQDK---IK 81

Db 10 NKQGTGTLGAGAGALLGSQFGKGGQGLVGVGALLAVLGQIGAGMDQDRRLAELT 69

QY 82 LNSLEKVVAGQVTRWNPDTGNSYSVEPVRTYQRYNKQERRQCYREFQOKAMTAGOKO 141

Db 70 SORALEATPSGTSVEMRPDNGHGYVTPNKTY-----RNSTGYCREYQTVVIGGKQO 124

QY 142 ELYGTAC 148

Db 125 KAYGNAC 131

## RESULT 11

Q52637

AC Q52637 PRELIMINARY; PRT; 131 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 17 KDA ANTIGEN (FRAGMENT)

OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia

NCBI\_TaxID=789;

RP SEQUENCE FROM N.A.

RA Werren J.H., Hurst G.D., Zhang W., Breuer J.A., Stouthamer R.,

RT "Rickettsial relative associated with male killing in the ladybird

RL J. Bacteriol. 176:388-394(1994).

FT EMBL: U04162; AAA19235.1; -

NON\_TER 131

SEQUENCE 131 AA: 13344 MW; A1DCF71050DF52DF CRC64;

Query Match

Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;

Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGATGAVGAGVAGQLFGKSGRVAMAIGAVLGLIGSKIGQSDQDK---IK 81

Db 10 NKQGTGTLGAGAGALLGSQFGKGGQGLVGVGALLAVLGQIGAGMDQDRRLAELT 69

QY 82 LNSLEKVVAGQVTRWNPDTGNSYSVEPVRTYQRYNKQERRQCYREFQOKAMTAGOKO 141

Db 70 SORALEATPSGTSVEMRPDNGHGYVTPNKTY-----RNSTGYCREYQTVVIGGKQO 124

QY 142 ELYGTAC 148

Db 125 KAYGNAC 131

## RESULT 12

Q9L522

AC Q9L522 PRELIMINARY; PRT; 131 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE 17 KDA SURFACE ANTIGEN (FRAGMENT)

OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia

NCBI\_TaxID=47589;

RP SEQUENCE FROM N.A.

RA Simer J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;

FT MEDLINE=21091941; PubMed=11157215;

NON\_TER 131

SEQUENCE 131 AA: 13344 MW; A1DCF71050DF52DF CRC64;

Query Match

Best Local Similarity 28.4%; Score 237.5; DB 2; Length 131;

Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

RT "Isolation of a spotted fever group rickettsia, Rickettsia peacockii,  
in a Rocky Mountain wood tick, Dermacentor andersoni, cell line."  
RL Appl. Environ. Microbiol. 67:546-552(2001).  
DR EMBL: AF260571; AAF69012.1; -.

FT NON\_TER 1  
RC STRAIN-MAFE303099;  
SQ SEQUENCE 131 AA; 13413 MW; 228C020550CA9D0 CRC64;

Query Match 28.3%; Score 236.5; DB 2; Length 131;  
Best Local Similarity 37.8%; Pred. No. 6.6e-14;  
Matches 48; Conservative 22; Mismatches 48; Indels 9; Gaps 2;

QY 26 SROEVGAATGAVGVAGVAGLFGKSGRVAMAIGGAVGLGKSGKSGMDQDK----IK 81  
DB 10 NKGTGTLGAGAGALLGSGFGKSGKGLGVGVALLGAVLGIGAGMDQDRRLAELT 69  
QY 82 LNSLEKVKAGQVTRMNPPTGNSYSVEPVTRTYRKQERRQOYCFEFOOKAMIAQOKO 141  
DB 70 SORALETPAGSGSNVEMNPNGNYVTVPNKTY-----RNSTGQYCREYTOTVIGGKQO 124  
QY 142 EITGTAC 148  
DB 125 KAIGNAC 131

RESULT 13  
031208  
ID 031208 PRELIMINARY; PRT; 105 AA.

AC 031208;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE 17 KDA ANTIGEN (FRAGMENT).  
OS Rickettsia sp. "La Copita".  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=69475;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LA COPITA;  
RA MEDLINE=98367252; PubMed=9701930;  
RA Billings A.N., Yu X.J., Teel P.D., Walker D.H.;  
RT "Detection of a spotted fever group rickettsia in Amblyomma cajennense  
[Acari: Ixodidae] in south Texas".  
RL J. Med. Entomol. 35:474-478(1998).  
DR EMBL: AF03499; AAB86943.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 105 AA; 11236 MW; 7BD035104701D4FB CRC64;

Query Match 26.7%; Score 223.5; DB 2; Length 105;  
Best Local Similarity 41.2%; Pred. No. 7.4e-13;  
Matches 47; Conservative 17; Mismatches 33; Indels 17; Gaps 3;

QY 43 GOLFSGSGRVAMAIGGAVGLGKSGKSGMDQDK----IKLNSLEKVKAGQVTRM 98  
DB 3 GOLFVGV-----GALLGAVLGIGAGMDQDRRLAELTQSRLETPAGSGSNVEM 54  
QY 99 NPDTGNSYSVEPVTRTYRKQERRQOYCFEFOOKAMIAQOKOYEITGTACROPD 152  
DB 55 NPDGNGYVTPNKTY-----RNSTGQYCREYTOTVIGGKQKAYGNACRQPD 103

RESULT 14  
ID 0985G4 PRELIMINARY; PRT; 199 AA.  
AC 0985G4;  
DT 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
DE MUR7687 PROTEIN.  
GN Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAFE303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti".  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AF03012; BAB54098.1; -.  
KW Complete proteome.  
SQ SEQUENCE 199 AA; 20517 MW; 98E6F20A734637AA CRC64;

Query Match 17.1%; Score 143; DB 16; Length 199;  
Best Local Similarity 31.8%; Pred. No. 2.9e-05;  
Matches 35; Conservative 15; Mismatches 50; Indels 10; Gaps 3;

QY 50 SGRVAMAIGGAVLGKSGKSGMDQDKIKL----NLSLEKVKAGQVTRMNPDTGNS 105  
DB 92 SGKVTKSLISAMDGLGSGISGNLSDEKRSALAEYKALEYTTSGQKVAWKDQASHY 151  
QY 106 YSEVPVTRTYRKQERRQOYCFEFOOKAMIAQOKOYEITGTACROPGRW 155  
DB 152 GEVVPAPY-RVGSQD-----CROYTOTVFTGGAGVATARGACHNADGWS 195

RESULT 15  
09AGC7  
ID 09AGC7 PRELIMINARY; PRT; 77 AA.

AC 09AGC7;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE 17 KDA ANTIGEN (FRAGMENT).  
OS Rickettsia typhi.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=785;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Paspaliaris V., Liedtke B., Viletta L., Whiting J.L.;  
RT "Rickettsia typhus and Rickettsia felis rickettsioses in Australia."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF336794; AAK21272.1; -.

FT NON\_TER 1  
SQ SEQUENCE 77 AA; 7955 MW; B1E447C037263918 CRC64;

Query Match 16.7%; Score 140; DB 2; Length 77;  
Best Local Similarity 38.7%; Pred. No. 1.6e-05;  
Matches 29; Conservative 13; Mismatches 29; Indels 4; Gaps 1;

QY 43 GOLFSGSGRVAMAIGGAVGLGKSGKSGMDQDK----IKLNSLEKVKAGQVTRM 98  
DB 3 GSGFGHCKGLGVGVALLGAVLGIGAGMDQDRRLAELTQSRALLESAPSGSNVEM 62  
QY 99 NPDTGNSYSVEPVTRTYRKQERRQOYCFEFOOKAMIAQOKOYEITGTACROPD 113  
DB 63 NPDGNGYVTPNKTY-----RNSTGQYCREYTOTVIGGKQKAYGNACRQPD 77

Search completed: October 27, 2002, 11:10:22  
Job time : 19.1865 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:36 : Search time 6.15544 Seconds

(without alignments)  
1019.028 Million cell updates/sec

Title: US-09-677-374-2

Sequence: 1 MNRGCLQSSLIITSVLVG.....IYGTACRQDPGRKQVISTEK 162

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296	35.4	159	17KD_RICPR	P16624 rickettsia
2	294	35.2	159	17KD_RICPA	O52764 rickettsia
3	293	35.0	159	17KD_RICCN	P05372 rickettsia
4	284.5	34.0	159	17KD_RICTY	P22882 rickettsia
5	276	33.0	154	17KD_RICAU	P50928 rickettsia
6	263	31.5	154	17KD_RICAM	P50927 rickettsia
7	261	31.2	154	17KD_RICPA	P50931 rickettsia
8	261	31.2	154	17KD_RICRH	P50929 rickettsia
9	258	30.9	154	17KD_RICMO	P29687 rickettsia
10	157	18.8	80	17KD_RICCA	P31484 yersinia en
11	113.5	13.6	155	PCP_YEREN	P53741 yersinia en
12	105.5	12.6	155	SLVB_ECOLI	O53549 salmonella
13	105.5	12.6	155	SLVB_SALT	P10335 haemophilus
14	102	12.2	155	PCP_HAEN	P37796 escherichia
15	99	11.8	179	YCFJ_ECOLI	P06384 bos taurus
16	90	10.8	526	KICJ_BOVIN	P54334 bacillus su
17	90	10.8	1332	XKDO_BACSU	P45931 bacillus su
18	89	10.6	1385	YOBQ_BACSU	P37723 salmonella
19	87.5	10.5	72	OSMB_SALT	P17853 salmonella
20	86.5	10.3	72	OSMB_ECOLI	P17853 salmonella
21	84.5	10.1	243	CYSH_SALT	O92188 ratius norv
22	84.5	10.1	301	STXG_RAT	O00310 candida alb
23	84.5	10.1	431	KRE2_CANAL	P33476 bluetongue
24	84.5	10.1	526	VP5_BTV11	O6867 aquilex aeo
25	82	9.8	132	V615_AOUAE	P13135 bos taurus
26	82	9.8	263	CANS_BOVIN	P48837 saccharomyc
27	81.5	9.7	541	NUS7_YEAST	P19837 nephila cla
28	81.5	9.7	747	SPD1_NEPCL	P46724 mycobacteri
29	80.5	9.6	467	HEM1_MYCLE	P34309 caenorhabdi
30	79.5	9.5	113	YKR3_CABEL	P06813 oryctolagus
31	79.5	9.5	266	CANS_RABIT	P19482 bos taurus
32	79.5	9.5	359	ATPA_BOVIN	P15999 ratius norv
33	79.5	9.5	543	ATPA_RAT	

## ALIGNMENTS

RESULT 1	ID	17KD_RICPR	STANDARD:	PRT:	159 AA.
AC	P16624	17KD_RICPR			
DT	01-AUG-1990 (Rel. 15, Created)				
DT	01-AUG-1991 (Rel. 19, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	17 kDa surface antigen precursor.				
GN	OMP OR RP833.				
OS	Rickettsia prowazekii.				
OC	Bacteria: Proteobacteria; alpha subdivision: Rickettsiales;				
OC	Rickettsiaceae; Rickettsiaseae; Rickettsia.				
OX	NCBI_Taxid=782;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MADRID E.				
RX	MEDLINE=69359171; Pubmed=2768201;				
RA	Anderson B.E., Tzianabos T.;				
RT	"Comparative sequence analysis of a genus-common rickettsial antigen				
RT	gene.";				
RL	J. Bacteriol. 171:5199-5201(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MADRID E.				
RX	MEDLINE=99039499; Pubmed=982893;				
RA	Andersson S.G.E., Zomrodipour A., Andersson J.O.,				
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,				
RT	"The genome sequence of Rickettsia prowazekii and the origin of				
RT	mitochondria.";				
RL	Nature 396:133-140(1998).				
CC	-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid				
CC	anchor (Probable).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M28482; AAA26378.1; ALT_SEQ.				
DR	EMBL; AJ235273; CAA15258.1; -				
DR	PIR; D33971; D33971.				
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.				
KW	Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.				
FT	SIGNAL	1	19		
FT	CHAIN	20	159	17 KDA SURFACE ANTIGEN.	
FT	LIPID	20	20	N-ACYL DIGLYCERIDE (PROBABLE).	
SQ	SEQUENCE	159 AA;	16672 MM;	A33D04B65EBB071 CRC64;	
Query Match		35.4%;	Score 296;	DB 1;	Length 159;
Best Local Similarity		38.1%;	Pred. NO. 9.9e-19;		
Matches	61;	Conservative	30;	Mismatches	55;
				Indels	14;
				Gaps	4;

```

QY 9 SLLIIISV---FLVGC--AQNFSROEVGATGAVVGAGVQLFGKSGSRVAMAIIGAVALG 63
DB 5 SKIMIIALASMLACNCGSGMANKOGTGLLGAGALLGSGFQGGGLVGVGVALLG 64
QY 64 GLISKIGSGMDQDK-----IKLNQSLKVKAGVOTRRNPDGNSYSVEPVRYQRYNK 119
DB 65 AVLGQIGASMDQDRRLLELTQSRALESNSNINRNDNNGYTPNKTY----- 119
QY 120 QERROOYCRFEQOKAMIAQOKOIEYTGACROPDGRMOWIS 159
DB 120 RNSAGOCREYTOTVIGGKQOKTYGACROPDGRMOWYVN 159

```

```

RESULT 2
17KD_RICJA
ID 17KD_RICJA STANDARD; PRT; 159 AA.
AC Q52764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
ON NCBI_TaxID=35790;
RP SEQUENCE FROM N.A.
RC STRAIN-YH.
RX MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
RL J. Clin. Microbiol. 33:487-489(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D16515; BAA03965.1;
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane: Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16554 MW; CDDCE7CBDDCD641 CRC64;

```

```

Query Match 35.2%; Score 294; DB 1; Length 159;
Best Local Similarity 38.1%; Pred. No. 1.5e-18;
Matches 61; Conservative 30; Mismatches 55; Indels 14; Gaps 4;
QY 9 SLLIIISV---FLVGC--AQNFSROEVGATGAVVGAGVQLFGKSGSRVAMAIIGAVALG 63
DB 5 SKIMIIALASMLACNCGSGMANKOGTGLLGAGALLGSGFQGGGLVGVGVALLG 64
QY 64 GLISKIGSGMDQDK-----IKLNQSLKVKAGVOTRRNPDGNSYSVEPVRYQRYNK 119
DB 65 AVLGQIGASMDQDRRLLELTQSRALESNSNINRNDNNGYTPNKTY----- 119
QY 120 QERROOYCRFEQOKAMIAQOKOIEYTGACROPDGRMOWIS 159
DB 120 RNSAGOCREYTOTVIGGKQOKTYGACROPDGRMOWYVN 159

```

```

RESULT 3
17KD_RICCN
ID 17KD_RICCN STANDARD; PRT; 159 AA.

```

```

AC P05372;
DT 01-NOV-1989 (Rel. 09, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RCI287.
OS Rickettsia conorii, and
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
ON NCBI_TaxID=781, 783;
RP SEQUENCE FROM N.A.
RC SPECIES-R.conorii, and R.rickettsii;
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RL gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-R.conorii; STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-R.rickettsii;
RX MEDLINE=87222152; PubMed=3108232;
RA Anderson B.E., Kennedy R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA Fu Z.Y., Bellini W.J.;
RT "Sequence analysis of the 17-kilodalton antigen gene from Rickettsia
RL rickettsii.";
RL J. Bacteriol. 169:2385-2390(1987).
RN [4]
RP SEQUENCE OF 1-30 FROM N.A.
RC SPECIES-R.rickettsii;
RX MEDLINE=89008059; PubMed=3139629;
RA Anderson B.E., Baumstark B.R., Bellini W.J.;
RT "Expression of the gene encoding the 17-kilodalton antigen from
RL Rickettsia rickettsii: transcription and posttranslational
RL J. Bacteriol. 170:4493-4500(1988).
RN [5]
RP -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
RP anchor (Probable).
RP -----
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
RP -----
DR EMBL; M28479; AAA63379.1;
DR EMBL; M28480; AAA26376.1;
DR EMBL; AE008675; AAL03825.1;
DR EMBL; M16486; AAA26381.1;
DR EMBL; J03371; -; NON_ANNOTATED_CDS.
DR PIR; A25972; A25972.
DR PIR; A31836; A31836.
DR PIR; A33971; A33971.
DR PIR; B33971; B33971.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Outer membrane: Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT CONFLICT 146 146 N->D (IN REF. 3).
FT CONFLICT 153 153 G->E (IN REF. 3).

```



```

SQ SEQUENCE 159 AA; 16581 MW; 206A2BBF74FCE169 CRC64;
Query Match 35.0%; Score 293; DB 1; Length 159;
Best Local Similarity 38.1%; Pred. No. 1.8e-18;
Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;
OY 9 SLLIITISV---FLVGC--AQNFSEVGAATGAVGVAGOLFGKSGSRVMAIGAVLG 63
   |::||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5 SKIMIILASMLQACNCPGSMNKGCTGLLGAGGALLGSQFGKGGQLVGVGVALLG 64
   ::||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 64 GLIGSKIGSMDQDK---IKLNSLEKYKAGVTRMRNPDGNSYSVEPRTYQRYNK 119
   ::||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 65 AVLGQIGAGDEQDRRLAELTSORALETPAPSGSNVEMRPNDCNGYVTPNKTY----- 119
   ::||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 120 QERRQOYCREFOQKAMTAGOKOETGTACRQPDGRNOVIS 159
   |::||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 120 RNSTGYCREYTOTVIGKQOKAYGNACRQPDGQWQVYN 159
   |::||::| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
17KD_RICAU STANDARD; PRT; 159 AA.
AC P22882;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
   gene."
RT J. Bacteriol. 171:5199-5201(1989).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
   anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
```

RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 CC Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 CC -i- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: U11013; AAB07704.1;  
 KM PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.  
 FT SIGNAL 1 19  
 FT CHAIN 1 19  
 FT LIPID 20 >154 BY SIMILARITY  
 FT NON\_TER 154 20 17 KDA SURFACE ANTIGEN.  
 SQ SEQUENCE 154 AA; 15879 MW; EAFBEAC29D943581 CRC64;  
 Query Match 31.5%; Score 263; DB 1; Length 154;  
 Best Local Similarity 36.8%; Pred. No. 6.4e-16;  
 Matches 57; Conservative 27; Mismatches 57; Indels 14; Gaps 4;  
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGRVAMAIGAVLG 63  
 DB 5 SKIMIALAASMLQACNGPGMKNKGCTGLLGAGGALLSGQFGKQOLGVGVGALLG 64  
 QY 64 GLIGSKTIGSDQDQDK---IKLNSLEKVKAGQVTRMRNPDGNSYSVEPVRYQRYNK 119  
 DB 65 AVLGQIGAGMDQDRIELTSQALLETAPSGSVNWRNPDNGNGVYTPNKTY----- 119  
 QY 120 QERROQYCRFEQOKAMIAQKQEIYGTACROPDGR 154  
 DB 120 RNSTGQYCREYQTQVYVIGGKQAKAYGNACLOPDQ 154

RESULT 7  
 ID 17KD-RICPA STANDARD; PRT; 154 AA.  
 AC P50930;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia parkeri.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_Taxid=35792;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-MACULATUM;  
 RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 CC -i- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: U17008; AAB82040.1;  
 KM PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.  
 FT SIGNAL 1 19  
 FT CHAIN 1 19  
 FT LIPID 20 >154 BY SIMILARITY  
 FT NON\_TER 154 20 17 KDA SURFACE ANTIGEN.  
 SQ SEQUENCE 154 AA; 15895 MW; OCF85AD5D96DFEBC CRC64;  
 Query Match 31.2%; Score 261; DB 1; Length 154;  
 Best Local Similarity 36.1%; Pred. No. 9.5e-16;  
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;  
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGRVAMAIGAVLG 63  
 DB 5 SKIMIALAASMLQACNGPGMKNKGCTGLLGAGGALLSGQFGKQOLGVGVGALLG 64  
 QY 64 GLIGSKTIGSDQDQDK---IKLNSLEKVKAGQVTRMRNPDGNSYSVEPVRYQRYNK 119  
 DB 65 AVLGQIGAGMDQDRIELTSQALLETAPSGSVNWRNPDNGNGVYTPNKTY----- 119  
 QY 120 QERROQYCRFEQOKAMIAQKQEIYGTACROPDGR 154  
 DB 120 RNSTGQYCREYQTQVYVIGGKQAKAYGNACLOPDQ 154

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT NON\_TER 154 154  
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;  
 Query Match 31.2%; Score 261; DB 1; Length 154;  
 Best Local Similarity 36.1%; Pred. No. 9.5e-16;  
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;  
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGRVAMAIGAVLG 63  
 DB 5 SKIMIALAASMLQACNGPGMKNKGCTGLLGAGGALLSGQFGKQOLGVGVGALLG 64  
 QY 64 GLIGSKTIGSDQDQDK---IKLNSLEKVKAGQVTRMRNPDGNSYSVEPVRYQRYNK 119  
 DB 65 AVLGQIGAGMDQDRIELTSQALLETAPSGSVNWRNPDNGNGVYTPNKTY----- 119  
 QY 120 QERROQYCRFEQOKAMIAQKQEIYGTACROPDGR 154  
 DB 120 RNSTGQYCREYQTQVYVIGGKQAKAYGNACLOPDQ 154

RESULT 8  
 ID 17KD-RICRH STANDARD; PRT; 154 AA.  
 AC P50931;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia rhipicephali.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_Taxid=33992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.  
 CC -i- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: U11020; AAB07706.1;  
 KM PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.  
 FT SIGNAL 1 19  
 FT CHAIN 1 19  
 FT LIPID 20 >154 BY SIMILARITY  
 FT NON\_TER 154 20 17 KDA SURFACE ANTIGEN.  
 SQ SEQUENCE 154 AA; 15895 MW; OCF85AD5D96DFEBC CRC64;  
 Query Match 31.2%; Score 261; DB 1; Length 154;  
 Best Local Similarity 36.1%; Pred. No. 9.5e-16;  
 Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;  
 QY 9 SLLITISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFGKSGRVAMAIGAVLG 63  
 DB 5 SKIMIALAASMLQACNGPGMKNKGCTGLLGAGGALLSGQFGKQOLGVGVGALLG 64  
 QY 64 GLIGSKTIGSDQDQDK---IKLNSLEKVKAGQVTRMRNPDGNSYSVEPVRYQRYNK 119  
 DB 65 AVLGQIGAGMDQDRIELTSQALLETAPSGSVNWRNPDNGNGVYTPNKTY----- 119  
 QY 120 QERROQYCRFEQOKAMIAQKQEIYGTACROPDGR 154  
 DB 120 RNSTGQYCREYQTQVYVIGGKQAKAYGNACLOPDQ 154

```

RESULT 9
17KD_RICMO STANDARD; PRT; 154 AA.
ID 17KD_RICMO
AC P50929;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OHIO 83-441;
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBD databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U11017; AAB07705.1; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E33DA CRC64;

Query Match 30.9%; Score 258; DB 1; Length 154;
Best Local Similarity 36.1%; Pred. No. 1.7e-15;
Matches 56; Conservative 28; Mismatches 57; Indels 14; Gaps 4;

OY 9 SLLIIVY---FLVGC--AQNFSRQEVGAATGAVGVAGOLFPGSGRYAMAIIGAVLG 63
DB 5 SKIMITIALASMLQACNGPGMKNQGTGLGAGAGALLSOFQGLVGVGVGALLG 64
OY 64 GLIGSKIGOSMDQDK---IKLNQSLKVKAGOVTRMRNDPTGNSYSVEPVRTYRYNK 119
DB 65 AVLGGQIGAGMDEQDRRLAELTTSORALETPASGSNVEMRNDPNCNGYVTPNKTY----- 119

OY 120 QERRQYGRREFQOKAMIAQKQELTYGTACROPDGR 154
DB 120 RNSGTGYCREYGTQYVVGKQOKAYGNACLPDGO 154

RESULT 10
17KD_RICCA STANDARD; PRT; 80 AA.
ID 17KD_RICCA
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=92108069; PubMed=1729713;
RX

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M82879; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR000437; PROK_LIPOPROT.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 157; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 4e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

OY 43 GOLFPGSGRYAMAIIGAVLGIGLSKIGOSMDQDK---IKLNQSLKVKAGOVTRMR 98
DB 1 GSGFGKKGGLIGVACALLGAILGNQIGAGMDEQDRRLAELTTSORALETPGSTSIEMR 60

OY 99 NPDTGNSYSVEPVRTYQ 115
DB 61 NPDGNGYGYVTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD; PRT; 155 AA.
ID PCP_YEREN
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcg precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RX MEDLINE=9211089; PubMed=1732192;
RA Baeninger A.J., Hanke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli.";
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -----
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURIO SLI7B AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60448; CAA42977.1; -
DR

```

```

DR PIR: S23787; S23787.
DR HSSP: P00778; 1P04.
KW PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
FT Outer membrane: Lipoprotein; Signal.
FT CHAIN 1 17
FT LIPID 18 155
FT SEQUENCE 155 AA: 15362 MW: 8406B2132E849FA CRC64;
SQ
Query Match
Best Local Similarity 24.7%; Score 113.5; DB 1; Length 155;
Matches 33; Conservative 19; Mismatches 43; Indels 51; Gaps 4;
.OY 10 SLIISVFLVGCAGN-----FSRQ-----VGAT 34
Db 7 AVAIAVATLTCGANNNTLSGDVFSASQAKOVITYGTILSVRPVTTGCGDNNMGAIG 66
OY 35 GAVYGVGVGQLFGKSGRVNAMAIGAVLGLISKIGSQMDQDKIKL----- 82
Db 67 GAVLGGFLGNTVGGGTGRSLATAGAVAGSGAGGQVQAGMNTDGVLEVRKDDGTTILV 126
OY 83 --NLSLEKVKAGQVTRWRNPDTGNSYSVP 110
Db 127 VQKQGPTRFSVQ--RVMLASSGTVTVSP 154

RESULT 12
SLIB_ECOLI
ID SLIB_ECOLI STANDARD: PRT: 155 AA.
AC P55741; P76183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Outer membrane lipoprotein slib precursor.
GN SLIB OR B1641 OR 42655 OR ECS2350.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RA Ludwig A., Tengell C., Bauer S., Bubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.; Pubmed=8544813;
RT haemolytic and pore-forming protein in Escherichia coli.;
RL Mol. Gen. Genet. 249:474-486(1993).
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Mau B., Shao Y.; Pubmed=9278503;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Aliba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kaba H., Kashimoto K., Kimura S., Kitakawa M.,
RA Motomura K., Makino K., Miki T., Mizobuchi K., Morita H., Mori T.,
RA Oshima T., Saito N., Nakamura Y., Nishimoto H., Nishio Y.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiiuchi T.; Pubmed=9097039;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).

```

---

```

RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; Pubmed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Grofbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomous K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.; Pubmed=11258796;
RT "Genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509552;
RX MEDLINE=21156231; Pubmed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuka E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RP SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (potential).
CC - SIMILARITY: TO S.TYPHIMORIUM SLIB, H INFLUENZAE PCP AND
CC Y. ENTEROCOLITICA PCP.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC use by non-profit institutions. There are no restrictions on its
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000259; AAC74713.1;
CC DR EMBL: D90807; BAAL5402.1;
CC DR EMBL: AE005387; AAG56630.1;
CC DR EMBL: AP002558; BAB35773.1;
CC DR Ecogene; EG31409; SLIB
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Outer membrane: Lipoprotein; Signal; Complete proteome.
CC FT SIGNAL 1 17
CC FT CHAIN 18 155
CC FT LIPID 18 155
CC FT CONFLICT 98 18 N-ACYL DIGLYCERIDE.
CC SQ SEQUENCE 155 AA: 15602 MW: 543EB84069A5FA3 CRC64;
Query Match
Best Local Similarity 22.8%; Score 105.5; DB 1; Length 155;
Matches 33; Conservative 21; Mismatches 42; Indels 49; Gaps 4;
.OY 13 IISVFLVGCAGN-----FSRQ-----VGATGAV 37
Db 10 MVLGLVGVNNTLSGDVYTSKAKQVNVSYGTIVNRPVTTGCGDNNMGAIGAV 69
OY 38 VGVAGQLFGKSGRVNAMAIGAVLGLISKIGSQMDQDKIKLNSLEKVKAGQVTRWR 97
Db 70 LGFLGNTVGGGTGRSLATAGAVAGSGAGGQVQAGMNTDGVLEVRKDDGTTILV 126
OY 98 RNDTGNYSYVEPVTVYRKNQER 122
Db 119 ---DDGNTIMVQKGMTRFSRQ 140

RESULT 13
SLIB_SALTY
ID SLIB_SALTY STANDARD: PRT: 155 AA.
AC Q53549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane lipoprotein slib precursor.

```

```

GN SLVB OR STM1445 OR STY1677.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=96133688; PubMed=8544813;
RA Ludwig A., Tengel C., Bauer S., Bubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.;
RT "SLV, a regulatory protein from Salmonella typhimurium, induces a
RT haemolytic and pore-forming protein in Escherichia coli.";
RL Mol. Genet. 249:474-486(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor J.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Potential).
CC -I- SIMILARITY: TO E.COLI SLVB, H.INFLUENZA PCP AND Y.ENTEROCOLITICA
CC PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S80790; AAB35871.2; -
CC DR EMBL: AE008762; AAL20367.1; -
CC DR EMBL: AL627271; CAD01922.1; -
CC DR StyGene; SG10573; slvb.
CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Outer membrane; Lipoprotein; Signal; Complete proteome.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLVB.
CC FT LIPID 18 18 N-ACYL DIGLYCERIDE.
CC SQ SEQUENCE 155 AA; 15548 MW; 82FDDCCBABD5A7 CRC64;

```

```

Query Match 12.6%; Score 105.5; DB 1; Length 155;
Best Local Similarity 24.6%; Pred. No. 0.02;
Matches 34; Conservative 19; Mismatches 52; Indels 33; Gaps 3;

```

```

OY 4 GCGGSSSL-----ITISVFLVGCAGNFSRQEVGAATGAVGCGVACQ 44
DB 17 GCNNNDLSGVDYTASEAQOVNVTYGTIVNRPVQIQGGDDSNVIGATGAVLGFLGN 76

```

```

OY 45 LFKGSGRVAMALGAVLGLGSKIGOSMDQODKIKLNLSLEKYKAGVTRMRNPDTCN 104
DB 77 TIGGTRGRSLATRAAGAVAGGAGGQVQSAAMNKIQVEL--EIRK-----DDGN 122
OY 105 SYSVEPRTYQRYNKOER 122
DB 123 TIMVVKQGNTRFSAGQR 140

```

```

RESULT 14
PCP_HAEIN STANDARD; PRT; 155 AA.
ID PCP_HAEIN
AC P10325;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL
DE cross-reacting lipoprotein).
GN PCP OR LPP OR H11579.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88115138; PubMed=2828309;
RA Delch R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
RT outer membrane lipoprotein and an antigenically related 15,000-dalton
RT protein from Haemophilus influenzae.";
RL J. Bacteriol. 170:489-498(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -I- SIMILARITY: TO E.COLI AND S.TYPHIMURIUM SLVB AND TO
CC Y.ENTEROCOLITICA PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M18877; AAA24938.1; -
CC DR EMBL: U32832; AAC23228.1; -
CC DR PIR; B28543; B28543.
CC DR TIGR; H11579; -
CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Outer membrane; Lipoprotein; Signal; Complete proteome.
CC FT SIGNAL 1 18
CC FT CHAIN 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.
CC FT LIPID 19 19 N-ACYL DIGLYCERIDE.
CC FT LIPID 15 143 CSVAEVEF -> VAGRRVRI (IN REF. 1).
CC SQ SEQUENCE 155 AA; 15425 MW; D7880327FCF0C985 CRC64;

```

```

DR   Ecogene_EG12444; ycfJ.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM 5
SQ   SEQUENCE 179 AA; 18920 MW; BA5EB0B56D45609 CRC64;
      POTENTIAL.
Query Match      11.8% Score 99; DB 1; Length 179;
Best Local Similarity 27.1% Pident. No. 0.082;
Matches 26; Conservative 12; Mismatches 32; Indels 26; Gaps
QY   31 GAATGAATGAAAGQLPFKSGSRVAAIMAGGAVLGLIGSKTIGSQMDDDKTKLNQSEKVK 90
      1: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   73 GSVLGAAGVATIGHDFGSGRGRKQKVAATVVAALGGYAGAQIGQSLQESD----- 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   91 AGQVTRWRNPDITGNSISVEPRTYQRYNKNQEPDQY 126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB   121 -----TYTTTQQRCKTVYDQKSEKMLGTY 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



C:Species: Rickettsia conorii  
 C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
 C:Accession: B33971  
 R:Anderson, B.E.; Tzianabos, T.  
 J. Bacteriol. 171, 5199-5201, 1989  
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
 A:Reference number: A33971; MUID:89359171  
 A:Accession: B33971  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <AND>  
 A:Cross-references: GB:M28479; NID:9152463; PIDN:AAA2637.1; PID:9152464  
 C:Superfamily: rickettsial common antigen

Query Match  
 Best Local Similarity 35.0%; Score 293; DB 2; Length 159;  
 Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;  
 QY 9 SSLIITSV---FLVGC--AONFSROEVGATGAVGVAGQLFGKSGRVAMAIIGAVIG 63  
 Db 5 SKIMTIALATSMLOACNGPGMKNKGTGTLGAGAGALLGSOFKKGQLVGVGALLG 63  
 QY 64 GLIGSKIGOSMDQDK----IKLNSLEKVKAGQVTRMNPDTGNSYSVEPVRTYQRYNK 119  
 Db 65 AVLGGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNPNGNGYVTPNKTY----- 119  
 QY 120 QERRQYCREFOOKAMTAGOKOELTYGTACROPDGRMOYIS 159  
 Db 120 RNSTGQYCRETYTQTVVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 3  
 A33971  
 Rickettsial common antigen precursor - Rickettsia rickettsii  
 C:Species: Rickettsia rickettsii  
 C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
 C:Accession: A33971  
 R:Anderson, B.E.; Tzianabos, T.  
 J. Bacteriol. 171, 5199-5201, 1989  
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
 A:Reference number: A33971; MUID:89359171  
 A:Accession: A33971  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <AND>  
 A:Cross-references: GB:M28480; NID:9152457; PIDN:AAA2637.1; PID:9152458  
 C:Superfamily: rickettsial common antigen

Query Match  
 Best Local Similarity 35.0%; Score 293; DB 2; Length 159;  
 Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;  
 QY 9 SSLIITSV---FLVGC--AONFSROEVGATGAVGVAGQLFGKSGRVAMAIIGAVIG 63  
 Db 5 SKIMTIALATSMLOACNGPGMKNKGTGTLGAGAGALLGSOFKKGQLVGVGALLG 63  
 QY 64 GLIGSKIGOSMDQDK----IKLNSLEKVKAGQVTRMNPDTGNSYSVEPVRTYQRYNK 119  
 Db 65 AVLGGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNPNGNGYVTPNKTY----- 119  
 QY 120 QERRQYCREFOOKAMTAGOKOELTYGTACROPDGRMOYIS 159  
 Db 120 RNSTGQYCRETYTQTVVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 4  
 G97860  
 17K surface antigen precursor [imported] - Rickettsia conorii (strain Malish 7)  
 C:Species: Rickettsia conorii  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
 C:Accession: G97860  
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
 science 293, 2033-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A:Reference number: A97700; MUID:21442074; PMID:11557893  
 A:Accession: G97860  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <KUR>  
 A:Cross-references: GB:AEO06914; PIDN:AAL03825.1; PID:915620425; GSPDB:GN00173  
 C:Genetics:  
 A:Gene: omp  
 C:Superfamily: rickettsial common antigen

Query Match  
 Best Local Similarity 35.0%; Score 293; DB 2; Length 159;  
 Matches 61; Conservative 29; Mismatches 56; Indels 14; Gaps 4;  
 QY 9 SSLIITSV---FLVGC--AONFSROEVGATGAVGVAGQLFGKSGRVAMAIIGAVIG 63  
 Db 5 SKIMTIALATSMLOACNGPGMKNKGTGTLGAGAGALLGSOFKKGQLVGVGALLG 63  
 QY 64 GLIGSKIGOSMDQDK----IKLNSLEKVKAGQVTRMNPDTGNSYSVEPVRTYQRYNK 119  
 Db 65 AVLGGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNPNGNGYVTPNKTY----- 119  
 QY 120 QERRQYCREFOOKAMTAGOKOELTYGTACROPDGRMOYIS 159  
 Db 120 RNSTGQYCRETYTQTVVIGKQOKAYGNACROPDGMQOVYN 159

RESULT 5  
 C33971  
 Rickettsial common antigen precursor - Rickettsia typhi  
 C:Species: Rickettsia typhi  
 C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
 C:Accession: C33971  
 R:Anderson, B.E.; Tzianabos, T.  
 J. Bacteriol. 171, 5199-5201, 1989  
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
 A:Reference number: A33971; MUID:89359171  
 A:Accession: C33971  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <AND>  
 A:Cross-references: GB:M28481; NID:9152459; PIDN:AAA26377.1; PID:9152460  
 C:Superfamily: rickettsial common antigen

Query Match  
 Best Local Similarity 34.0%; Score 284.5; DB 2; Length 159;  
 Matches 55; Conservative 25; Mismatches 49; Indels 9; Gaps 2;  
 QY 26 SROEVGATGAVGVAGQLFGKSGRVAMAIIGAVIGGLIGSKIGOSMDQDK----IK 81  
 Db 27 NKQGTGTLGAGAGALLGSOFKKGQLVGVGALLGAGLIGASLDQDKLLELT 86  
 QY 82 INQSLKVKAGQVTRMNPDTGNSYSVEPVRTYQRYNKQERRQYCREFOOKAMTAGOKO 141  
 Db 87 SORALESPAGSGSNVEMRNPNGNGYVTPNKTY----- RNSTGQYCRETYTQTVVIGKQO 141  
 QY 142 EIYGTACROPDGRMOYIS 159  
 Db 142 TTYGNACROPDGMQOVYN 159

RESULT 6  
 A25972  
 17K antigen precursor - Rickettsia rickettsii  
 C:Species: Rickettsia rickettsii  
 C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 20-Aug-1999  
 C:Accession: A25972  
 R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; Mcdade, J.E.; Fu, Z.Y.  
 J. Bacteriol. 169, 2385-2390, 1987  
 A:Title: Sequence analysis of the 17-kilodalton antigen gene from Rickettsia ricketts  
 A:Reference number: A25972; MUID:87222152





A:Residues: 1-155 <BAE>  
 A:Cross-references: EMBL:X60448; NID:948577; PIDN:CAA42977.1; PID:948579  
 C:Superfamily: PAL cross-reacting 11protein

## Query Match

Best Local Similarity 13.6%; Score 113.5; DB 2; Length 155;  
 Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

QY 10 SLIISVFLVCAQN-----FSROE-----  
 Db 7 AVALAATLTTCANNNTLSDVFSASQAKOVITVTGTLTSVPYITIGGDONNMKAGIG 34  
 QY 35 GAVVGVAGVQLFGKSGSRVAMAGAVLGLIGSKIGSGMDPODQIKL-----  
 Db 67 GAVLGGFLGNTVGGCTGRTSLATTAAGAVAGMAGGCGVAGMNRDGVOLLEVRKDDGTTILV 82  
 QY 83 --NOSLEKVKAGQVTRRRNPDGNSYSYEP 110  
 Db 127 VQKGPTRFSVGO--RVMGLASSGRTVTSP 154

## RESULT 11

B82837

conserved hypothetical protein XF0178 [Imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 20-Aug-2000  
 C:Accession: B82837  
 R:Anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; PMID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82837  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-257 <SIM>  
 A:Cross-references: GB:AE003872; GB:AE003849; NID:99104975; PIDN:AAF82991.1; GSPDB:GN001

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H.; Neto, E.; Docena, C.; El-Dorzy, H.; Pacincaul, A.P.; Ferreira, A.J.S.  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Krumm, E.E.; Laig, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak, M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0178

## Query Match

Best Local Similarity 13.4%; Score 112; DB 2; Length 257;  
 Matches 20; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 31 GAATGAVVGVAGVQLFGKSGSRVAMAGAVLGLIGSKI 70  
 Db 105 GTAICATLIGLVGNQFGHGNRKALTRAAGAVAGGFIGNEV 144

## RESULT 12

AD2696

11pA protein [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence-revision 11-Jan-2002 #text-change 11-Jan-2002  
 C:Accession: AD2696  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ka

ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AD2696  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-142 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AAL41986.1; PID:g17739358; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: 11pA  
 A:Map position: circular chromosome

Query Match  
 Best Local Similarity 13.3%; Score 111.5; DB 2; Length 142;  
 Matches 41; Conservative 21; Mismatches 60; Indels 29; Gaps 9;

QY 10 SLIISVFLVCAQNFSGEAGATGAVVGVAGVQLFGKSGSRVAMAGAVLGLIGSK 69  
 Db 12 SLICVSM-LSAC-----TTGTRPAG--GSLFGR-SAQSTPFLANLGGIVG-K 56  
 QY 70 IGOSDPODDQIKL-----NOSLEKVKAGQVTRRRNPDGNSYSYEPVRYRYNKKERRQ 124  
 Db 57 SGVELDGDQDTKALEEYKALETAPVGPVITGDDVAGVYANAP---YQVGN----- 107  
 QY 125 QYCNFQOKAMINAGQKQEIYGTACQPDGRW 155  
 Db 108 QNCRQYSHLTIVDGRDTRVRGACACRNDGWSW 138

## RESULT 13

D97478

11pA protein [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 30-Sep-2001 #sequence-revision 30-Sep-2001 #text-change 11-Jan-2002  
 C:Accession: D97478  
 R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
 A:Reference number: A97359; PMID:11743194  
 A:Accession: D97478  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-125 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK6781.1; PID:g15155981; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_1782  
 A:Map position: circular chromosome

Query Match  
 Best Local Similarity 13.3%; Score 111; DB 2; Length 125;  
 Matches 36; Conservative 18; Mismatches 61; Indels 16; Gaps 6;

QY 30 VCAATGAVVGVAGVQLFGKSGSRVAMAGAVLGLIGSKIGSGMDPODQIKL-----NOS 85  
 Db 2 LSACTTGTGTRPAGSLEGR-SAQSTPFLANLGGIVG-KSGVELDGDQDTKALEEYKA 59  
 QY 86 LEKVKAGQVTRRRNPDGNSYSYEPVRYRYNKKERRQYCNFQOKAMINAGQKQEIY 144  
 Db 60 LETAPVGTPIVITGDDVYKQVYANAP---YQVGN-----QNCRQYSHLTIVDGRDTRVR 110  
 QY 145 GTACRQPDGRW 155  
 Db 111 GAACRNDGWSW 121

## RESULT 14

S58234

11pA protein - Rhizobium leguminosarum  
 C:Species: Rhizobium leguminosarum



**THIS PAGE BLANK (USPTO)**



us-09-677-374-2.oligo.ra

RESULT 2  
US-09-247-806-1  
Sequence 1, Application US/09247806  
Patent No. 6280747  
GENERAL INFORMATION:  
APPLICANT: PHILIPPE, Michel  
APPLICANT: GARSON, Jean-Claude  
TITLE OF INVENTION: COSMETIC  
TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION  
TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT  
TITLE OF INVENTION: ANTI-C

Query Match	9.78;	Score 81.5;	DB 4;	Length 651;
Best Local Similarity	36.8%;	Pred. No. 0.8;		
Matches	25;	Conservative	3;	Mismatches 17

RESULT 3  
US-08-425-069-2  
Sequence 2, Application US/08425065  
Patent No. 5728810  
GENERAL INFORMATION

APPLICANT: Lewis, Randolph V.  
 APPLICANT: Xu, Ming  
 TITLE OF INVENTION: Himan, Michael B.  
 TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
 NUMBER OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
 NUMBER OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
 NUMBER OF SEQUENCES: 69  
 CORRESPONDENCE ADDRESS:

STREET: 301 No. 5728810th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,069  
FILING DATE: 19-Apr-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
1

Best Local Similarity 9.78; Score 81.5; DB 1; Length 718  
Matches 25; Conservative 3; Mismatches 17;

RESULT 4  
 S-08-317-844B-2  
 Sequence 2, Application US/08317844B  
 Patent No. 5989894  
 GENERATED BY

ADDITIONAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
APPLICANT: Himm, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
NUMBER OF INVENTION: CONTAINING THE ISOLATED DNA,  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESS:

SIREY: 301 No. 5989894th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046

COMPILED READABLE FORM:  
 MEDIA TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/17,844B  
 FILING DATE: 04-OCT-1994  
 CLASSIFICATION:

NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-105105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 241-1300  
TELEFAX: (703) 241-2848  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 718 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
-317-844B-2

Query Match 9.78; Score 81.5; DB 2; Length 718

RESULT 6  
US-07-803-633A-13

Patent No. 5369025  
GENERAL INFORMATION:  
APPLICANT: MAZERIAN, Keyvan

APPLICANT: YANAGIDA, No. 53690250ru  
APPLICANT: OGAWA, Ryohei  
APPLICANT: LI, Yi  
APPLICANT: YOUNG, JENNIFER ANN

NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:

ADDRESS: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 301 No. 5369025th Washington Street  
CITY: Falls Church

```

STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTED ADDRESS FROM

```

```

COMPTON READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: DOS/MS-DOS
?

```

```

1 SOFTWARE: Patentin Release #1.0, Version #1.25
2 CURRENT APPLICATION DATA:
3   APPLICATION NUMBER: US-07/R03.633A
4   APPLICATION NITMBER:

```

FILING DATE: 19911210  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1644-103P

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 241-1300  
TELEFAX: (703) 241-2848

TELEX: 248345  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
OFF

LENGTH: 603 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein

Query Match	9 34	Score 78	DB 1	Length 865
US-07-803-633A-13				
US-07-803-633A-13				

Best Local Similarity 25.4%; Pred. No. 2.9;  
Matches 36; Conservative 30; Mismatches 48; Indels 28; Gaps 7;

QY 17 FLVCAQNFESRQ - EVGCAATGAVVGAGQLPFKGSGRV-----AMAIIGCAVLGSLIG 67  
 688 FMNGLELFFNMGVGGAIGKVVVGAAGAIVSTISGSVAFMSNPFALAIIGLIITAGIAGA 747  
 bb

```
Oy      68 SKIGQSMDOQDKIKLN-----QSLEKVKACQVTRWRNPPDTGNSYSVEPVRTYORYNKO 120
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
```

Db	748	AFL-AYRYVNKLKSNPKMALKYPMITEVLKA-QATRELHGSEDD-----LERTSID	796
Oy	121	ERRRQYCRE-FQOKAMTAGOKO	141

```
Db      797  ERKLEAREMIKYMALVSAEER 818
```

[illegible]

Sequence ID: APPRIATION: 00/00000000  
Patent No. 6309861  
GENERAL INFORMATION:  
APPLICANT: Ambrosius, Dorothea

APPLICANT: Rudolph, Rainer  
APPLICANT: Schaeffner, Joerg  
APPLICANT: Schwarz, Elisabeth

; TITLE OF INVENTION: Process for the production of naturally folded and secreted p  
 ; FILE REFERENCE: Case 20379

CURRENT APPLICATION NUMBER: US/09/553,498  
CURRENT FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: EP99107412.1  
NUMBER OF SEQ ID NOS: 10  
SEQ ID NO: 8  
LENGTH: 255  
TYPE: PRN  
ORGANISM: E. coli  
US-09-553-498-8

Query Match  
Best Local Similarity 20.98; Score 77.5; DB 4; Length 255;  
Matches 37; Conservative 28; Mismatches 81; Indels 31; Gaps 7;

Db 6 GSSLIITSVFLVGCAGNFSROEVAATGAVY---GGVAQOLFSGSGVAMAGAV 61  
83 LGMTSLREPDYAMTCARDYGAY-WGCGTYYVSSGGGGGGGGSDIELTOSPAI 141  
Oy 62 LGGLGSKTGSDMDODKIK-LN-----QSLERKAGQVTRMRNPDGNSY 106  
142 MSASPGKRYMTCSASSSVRYMMWFOOKSGTSPKRWIYDTSKLSSGVPAFSGSGSTY 201  
Oy 107 VSEFVRYQRYNROEROOYCEFOOKMI--AGQKEITGTACRQPDGRWVISTE 161  
Db 202 SL-----TISMEADATYTCQWMSNPLTFGACTKLEKRAAEO-----KLISEE 249

RESULT 8  
US-09-067-351-2  
Sequence 2, Application US/09067351  
Patent No. 5994081  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: Baughn, Mariah  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,351  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-067-351-2

Query Match  
Best Local Similarity 22.44; Score 77; DB 2; Length 551;  
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

Db 8 GSSLIITSVFLVGCAGNFSROEVAATGAVVAGVAGOLFSGSGRVAMAGAVLGGLIG 67  
54 GASFGSRSLYNLGAKRVSLNCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110  
Oy 68 SK-----IGSMDODKIKLNQSLERKAGQVTRMRNPDGNSYSEFVRY 114  
Db 111 PSFVPCPGCIQEVTVNOSLPLHLQIDPTIORVAREERIKTLNNFTSFIDVRL 170  
Oy 115 QRYNK 119  
Db 171 EQONK 175

RESULT 9  
US-09-360-490-2  
Sequence 2, Application US/09360490  
Patent No. 6221843  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: Baughn, Mariah  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/360,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-360-490-2

Query Match  
Best Local Similarity 9.28; Score 77; DB 4; Length 551;  
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

Db 8 GSSLIITSVFLVGCAGNFSROEVAATGAVVAGVAGOLFSGSGRVAMAGAVLGGLIG 67  
54 GASFGSRSLYNLGAKRVSLNCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110



Qy 68 SK-----IGSQMDQDKIKLNQSLKRVKAGVTRMNPDTGNSYVEPVRTY 114  
Db 111 PSEFVCPGCGIQTAVTQSLTLPILHLOIDPIQVRAERBQKITLNNKFTSFIDKRVFL 170  
Qy 115 QRYNK 119  
Db 171 EQQNK 175

RESULT 10  
US-08-374-077C-2  
; Sequence 2, Application US/08374077C  
; Patent No. 6027912  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubald, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha  
; TITLE OF INVENTION: Calcium Channel Subunit  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374, 077C  
; FILING DATE: 19-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm M.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 022650-264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-374-077C-2

Query Match 9.2%; Score 76.5; DB 3; Length 2516;  
Best Local Similarity 24.8%; Pred. No. 17;  
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VGAATGAVGVAGQLFGKSGRYAMAIG--AVLGLISKIGQS---MDQDKIKLN 83  
Db 2389 IGSSNGSIFGSGAGLGAGSGGVG-GLGSSSTRNMFSGSGSPSLSPQHGYSGTLN 2447  
Qy 84 -----QSLKRVKAGVTRMNPDTGNSYVEPVRTYQRINKORROOYCREFOQKAMIA 137  
Db 2448 SPPIDNRRLRRVATVTTNNNNKSQVSONSSSLNVRANANSQNMSPGPOVQOQSPPLR 2507  
Qy 138 GQKOEIYGT 146  
Db 2508 GQGNQTYSS 2516

RESULT 11  
US-08-895-590-2  
; Sequence 2, Application US/08895590  
; Patent No. 6207410

; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubald, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,590  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374,888  
; FILING DATE: 19-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm M.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 022650-263  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;  
Best Local Similarity 24.8%; Pred. No. 17;  
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

Qy 30 VGAATGAVGVAGQLFGKSGRYAMAIG--AVLGLISKIGQS---MDQDKIKLN 83  
Db 2389 IGSSNGSIFGSGAGLGAGSGGVG-GLGSSSTRNMFSGSGSPSLSPQHGYSGTLN 2447  
Qy 84 -----QSLKRVKAGVTRMNPDTGNSYVEPVRTYQRINKORROOYCREFOQKAMIA 137  
Db 2448 SPPIDNRRLRRVATVTTNNNNKSQVSONSSSLNVRANANSQNMSPGPOVQOQSPPLR 2507  
Qy 138 GQKOEIYGT 146  
Db 2508 GQGNQTYSS 2516

RESULT 12  
US-07-910-760-12  
; Sequence 12, Application US/07910760  
; Patent No. 5683864  
; GENERAL INFORMATION:  
; APPLICANT: Houghton, Michael  
; APPLICANT: Choo, Qui-Lim  
; APPLICANT: Kuo, George  
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
; TITLE OF INVENTION: Antigens for use in Immunoassays for Anti-HCV Antibodies  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: P.O. Box 8097 (Int. Prop. R-440)  
; CITY: Emeryville

STATE: CA  
 COUNTRY: U.S.A.  
 ZIP: 94662-8097  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/910,760  
 FILING DATE: 07-JUL-1992  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Blackburn Esq., Robert P.  
 REGISTRATION NUMBER: 30,447  
 REFERENCE/DOCKET NUMBER: 0101.002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (510) 601-2702  
 TELEFAX: (510) 655-3542  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1021 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-07-910-760-12

Query Match  
 Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;  
 Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

OY 31 GAATGAVGVGACQLFGK-GSGRVMAI-----GCAVLGLIGSKI--GQSMDOQDKIKLN 83  
 DB 792 GAATAFAGAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLL 851  
 OY 84 QSL-----EKVKGQ-VTRWRN-----PDGNSYSVEPVRTYQRYNKO 120  
 DB 852 PATLSGALVGVCAAILRRHVGPBGAVQWMNRLIAFASRGNHVSPGNSSTNPKPOKK 911  
 OY 121 ERRO-----OYCREFOQKAMTAGCKOETIYTAC 148  
 DB 912 NKRNTNRPRDVKPPGGQIVGVYLLPRGRPLGVRAIRKTSERSOPRGRQPI--PKA 969  
 OY 149 ROPDGR-----WQYISTE 161  
 DB 970 RRPEGRFWAGPGYMPLYGNE 990

RESULT 13  
 US-08-440-519-12  
 Sequence 12, Application US/08440519  
 Patent No. 5712087  
 GENERAL INFORMATION:  
 APPLICANT: Houghton, Michael  
 APPLICANT: Choo, Qui-Lim  
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Chiron Corporation  
 STREET: P.O. Box 8097 (Int. Prop. R-440)  
 CITY: Emeryville  
 STATE: CA  
 COUNTRY: U.S.A.  
 ZIP: 94662-8097  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/440,519

FILING DATE: 12-MAY-1995  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/910,760  
 FILING DATE: 07-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Blackburn Esq., Robert P.  
 REGISTRATION NUMBER: 30,447  
 REFERENCE/DOCKET NUMBER: 0101.002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (510) 601-2702  
 TELEFAX: (510) 655-3542  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1021 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-440-519-12

Query Match  
 Best Local Similarity 9.1%; Score 76; DB 1; Length 1021;  
 Matches 42; Conservative 24; Mismatches 63; Indels 72; Gaps 9;

OY 31 GAATGAVGVGACQLFGK-GSGRVMAI-----GCAVLGLIGSKI--GQSMDOQDKIKLN 83  
 DB 792 GAATAFAGAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLL 851  
 OY 84 QSL-----EKVKGQ-VTRWRN-----PDGNSYSVEPVRTYQRYNKO 120  
 DB 852 PATLSGALVGVCAAILRRHVGPBGAVQWMNRLIAFASRGNHVSPGNSSTNPKPOKK 911  
 OY 121 ERRO-----OYCREFOQKAMTAGCKOETIYTAC 148  
 DB 912 NKRNTNRPRDVKPPGGQIVGVYLLPRGRPLGVRAIRKTSERSOPRGRQPI--PKA 969  
 OY 149 ROPDGR-----WQYISTE 161  
 DB 970 RRPEGRFWAGPGYMPLYGNE 990

RESULT 14  
 US-08-440-549-12  
 Sequence 12, Application US/08440549  
 Patent No. 6312889  
 GENERAL INFORMATION:  
 APPLICANT: Houghton, Michael  
 APPLICANT: Choo, Qui-Lim  
 TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Chiron Corporation  
 STREET: P.O. Box 8097 (Int. Prop. R-440)  
 CITY: Emeryville  
 STATE: CA  
 COUNTRY: U.S.A.  
 ZIP: 94662-8097  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/440,549  
 FILING DATE: 12-MAY-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/910,760  
 FILING DATE: 07-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Blackburn Esq., Robert P.



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:58:51 ; Search time 21.544 Seconds  
(without alignments)  
835.217 Million cell updates/sec

Title: US-09-677-374-2

Perfect score: 836  
Sequence: 1 NMRGLOGSLITISFLVG.....ITGTACRQPDGRMQVISTEK 162

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq-032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	836	100.0	162	22	AAG78025	Piscirickettsia
2	836	100.0	162	22	AAB81126	Ospa antigen amino
3	815	97.5	161	22	AAB81127	Optimised Ospa pro
4	815	97.5	256	22	AAB81128	C17E2 Ospa constru
5	112	13.4	20	22	AA881130	Ospa B-cell epitop
6	108	12.9	224	20	AA820105	Moraxella catarrha
7	105.5	12.6	223	20	AA734487	Porphomonas gling
8	105.5	12.6	230	20	AA734362	Porphomonas gling
9	102	12.2	154	11	AA805799	PBOHP-2 gene prod.
10	98	11.7	309	22	ABG15906	Novel human diagno
11	91	10.9	528	22	AA882611	Spider recombinant

12	86.5	10.3	2309	22	ABB6232	Drosophila melanog
13	82	9.8	666	22	ABB58019	Drosophila melanog
14	81.5	9.7	651	20	AA400697	Spider silk protei
15	81.5	9.7	718	12	AA14308	N.clavipes draglin
16	81.5	9.7	718	19	AAW53346	Nephila clavipes s
17	81.5	9.7	718	21	AAW59070	N. clavipes spider
18	80.5	9.6	102	22	AA441943	Human polypeptide
19	80.5	9.6	302	22	AA440157	Human polypeptide
20	80.5	9.6	542	22	ABB65790	Drosophila melanog
21	80.5	9.6	542	22	ABB65791	Drosophila melanog
22	80.5	9.6	542	22	ABB70501	Drosophila melanog
23	80	9.6	116	19	AA11028	H. pylori ORF O1cp
24	80	9.6	2017	22	ABG06301	Novel human diagno
25	80	9.6	2599	21	AA75098	Neisseria meningit
26	79	9.4	618	21	AA56803	Human prostate can
27	78	9.3	646	18	AAW27178	Nephila clavipes s
28	78	9.3	865	14	AAAR30169	Marek's Disease VI
29	77.5	9.3	255	21	AA11398	E. coli expression
30	77.5	9.3	255	22	AA74199	PelB-scFvOxazolom
31	77.5	9.3	255	22	AA70769	Expression plasmid
32	77.5	9.3	255	22	AA72020	E. carotovora PelB
33	77.5	9.3	900	22	ABG28648	Novel human diagno
34	77	9.2	147	22	ABB69847	Drosophila melanog
35	77	9.2	429	10	AA90064	Antigenic protein
36	77	9.2	551	21	AA52398	Human Keratin KERT
37	77	9.2	1251	22	ABB61254	Drosophila melanog
38	76.5	9.2	514	22	AAU36520	Pseudomonas derugi
39	76.5	9.2	2516	17	AAW01875	Neuronal Invertebr
40	76.5	9.2	2516	17	AAW01884	Invertebrate calci
41	76.5	9.2	2516	22	ABB61076	Drosophila melanog
42	76	9.1	300	21	AA805878	Merizite surface a
43	76	9.1	324	22	AA76751	Corynebacterium gl
44	76	9.1	346	22	AA692057	C glutamicum prote
45	76	9.1	346	22	AA76750	Corynebacterium gl

#### ALIGNMENTS

RESULT 1	
ID	AAG78025
AA	AAG78025 standard; Protein; 162 AA.
XX	
AC	AAG78025.
XX	
DT	15-JAN-2002 (first entry)
DE	Piscirickettsia salmonis polypeptide p10.6.
XX	
KW	Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
KW	septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;
KW	ATCC VR-1361.
XX	
OS	Piscirickettsia salmonis.
XX	
PN	WO200168865-A2.
XX	
PD	20-SEP-2001.
XX	
PF	12-MAR-2001; 2001WO-GH01055.
XX	
PR	11-MAR-2000; 2000GB-0005838.
PR	01-JUL-2000; 2000GB-0016080.
PR	01-JUL-2000; 2000GB-0016082.
XX	
XX	29-JUL-2000; 2000GB-0018599.
PA	(AQUA-) AQUA HEALTH EURO LTD.
XX	
PI	Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI	Burzio L;
XX	
DR	WPI: 2001-639050/73.
DR	N-PDB; AAH79040.

XX New nucleic acids encoding an amino acid sequence homologous to the  
 PT surface antigen present on Piscirickettsia salmonis are useful to  
 PT protect fish against piscirickettsiosis  
 XX Claim 6; Fig 5; 25pp; English.  
 PS  
 CC The invention relates to nucleic acid sequences and the encoded protein  
 CC of a least part of the surface antigen present on Piscirickettsia  
 CC salmonis for production of a vaccine with antibacterial activity to  
 CC protect fish against P. salmonis which causes piscirickettsiosis, also  
 CC known as salmonid rickettsial septicaemia.  
 XX Sequence 162 AA:  
 SO  
 Query Match  
 Best Local Similarity 100.0%; Score 836; DB 22; Length 162;  
 Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNRGCLGSSLLITISVFLVCAQNFPSROEYGAATGAVGVAGQLFGKSGRYAMAIGGA 60  
 Db 1 MNRGCLGSSLLITISVFLVCAQNFPSROEYGAATGAVGVAGQLFGKSGRYAMAIGGA 60  
 QY 61 VLGLIGSKIGSGMDQDKIKLNOSLEKVKAGOVTRMRNPDGNSYSVEPVRTYORYNKO 120  
 Db 61 VLGLIGSKIGSGMDQDKIKLNOSLEKVKAGOVTRMRNPDGNSYSVEPVRTYORYNKO 120  
 QY 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162  
 Db 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162  
 RESULT 2  
 ID AAB81126 standard; Protein; 162 AA.  
 XX AAB81126;  
 AC  
 DT 11-JUL-2001 (first entry)  
 XX  
 DE OSPA antigen amino acid sequence.  
 XX  
 KW Polkilothemic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
 XX SRS.  
 OS Piscirickettsia salmonis.  
 FH  
 FT Key Location/Qualifiers  
 FT Region 110..129  
 PN CA2281913-A1.  
 XX  
 PD 17-MAR-2001.  
 XX  
 PF 17-SEP-1999; 99CA-2281913.  
 XX  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 PA (KAYW/) KAY W W.  
 PA (BURI/) BURIAN J.  
 PA (KUZY/) KUZUK M A.  
 XX  
 PI Kay WW, Burian J, Kuzuk MA;  
 XX  
 DR WPI: 2001-316844/34.  
 DR N-PSDB; AAF86246.  
 XX  
 PT Method for protecting polkilothemic fish against salmonid rickettsial  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis  
 XX

PS Example 2; Fig 2b; 35pp; English.  
 XX  
 CC This invention relates to a method for the protection against infection  
 CC of a polkilothemic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkilothemic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents P. salmonis Ospa protein. An Ospa protein with an  
 CC N-terminal fusion partner is used in a vaccine to create an anti-Ospa  
 CC antibody response.  
 XX Sequence 162 AA:  
 SO  
 Query Match  
 Best Local Similarity 100.0%; Score 836; DB 22; Length 162;  
 Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNRGCLGSSLLITISVFLVCAQNFPSROEYGAATGAVGVAGQLFGKSGRYAMAIGGA 60  
 Db 1 MNRGCLGSSLLITISVFLVCAQNFPSROEYGAATGAVGVAGQLFGKSGRYAMAIGGA 60  
 QY 61 VLGLIGSKIGSGMDQDKIKLNOSLEKVKAGOVTRMRNPDGNSYSVEPVRTYORYNKO 120  
 Db 61 VLGLIGSKIGSGMDQDKIKLNOSLEKVKAGOVTRMRNPDGNSYSVEPVRTYORYNKO 120  
 QY 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162  
 Db 121 ERROQYCFREFOOKAMIAQOKOETIGTACROPDGRMVOISTEK 162  
 RESULT 3  
 ID AAB81127 standard; Protein; 161 AA.  
 XX AAB81127;  
 AC  
 DT 11-JUL-2001 (first entry)  
 XX  
 DE Optimised Ospa protein 17E2 amino acid sequence.  
 XX  
 KW Polkilothemic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
 XX SRS.  
 OS Piscirickettsia salmonis.  
 FH  
 FT Key Location/Qualifiers  
 FT Region 109..128  
 PN CA2281913-A1.  
 XX  
 PD 17-MAR-2001.  
 XX  
 PF 17-SEP-1999; 99CA-2281913.  
 XX  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 PA (KAYW/) KAY W W.  
 PA (BURI/) BURIAN J.  
 PA (KUZY/) KUZUK M A.  
 XX  
 PI Kay WW, Burian J, Kuzuk MA;  
 XX  
 DR WPI: 2001-316844/34.  
 DR N-PSDB; AAF86247.  
 XX  
 PT Method for protecting polkilothemic fish against salmonid rickettsial  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis  
 XX

PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -  
XX  
PS Disclosure; Fig 5; 35pp; English.  
CC This invention relates to a method for the protection against infection  
CC of a polkilothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC *salmonis*. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of a  
CC Ospa in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkilothermic fish, against the bacterial pathogen  
CC *P. salmonis*. The method is also useful for protecting against salmonid  
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
CC sequence represents optimised *P. salmonis* Ospa protein 17E2. The DNA  
CC encoding Ospa 17E2 (AA66247) has been optimised for expression in  
CC *Escherichia coli*. An Ospa protein with an N-terminal fusion partner is  
CC used in a vaccine to create an anti-Ospa antibody response.  
XX  
SQ Sequence 161 AA;  
  
Query Match 97.5%; Score 815; DB 22; Length 161;  
Best Local Similarity 98.8%; Pred. No. 5.1e-78;  
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
OY 3 RGCLQGSSLIITISVFLGCAQNFSEVGAATGAVGVAGOLFQKSGRYMAIGGAVL 62  
DB 2 RGCLQGSSLIITISVFLGCAQNFSEVGAATGAVGVAGOLFQKSGRYMAIGGAVL 61  
  
OY 63 GGLIGSKIGSMQDQDKIKLNQSLKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 122  
DB 62 GGLIGSKIGSMQDQDKIKLNQSLKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 121  
  
OY 123 RQCYCREFOOKAMTAGOKOETIYGTACPODGRWIVISTEK 162  
DB 122 RQCYCREFOOKAMTAGOKOETIYGTACPODGRWIVISTEK 161  
  
RESULT 4  
AAB81128  
ID AAB81128 standard; Protein: 256 AA.  
AC AAB81128;  
XX  
DT 11-JUL-2001 (first entry)  
XX  
DE C17E2 Ospa construct with N-terminal fusion partner.  
XX  
KW Polkilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;  
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
KW SRS; 17E2; fusion construct.  
XX  
OS *Piscirickettsia salmonis*.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 1..95  
FT /label= Undefined\_N-terminal\_fusion\_partner  
FT 96..256  
FT Region /label= C17E2\_Ospa  
FT /note= "Product of Ospa gene optimised for expression in  
FT *Escherichia coli*"  
XX  
XX CA2281913-A1.  
XX PN  
XX PD 17-MAR-2001.  
XX PF 17-SEP-1999; 99CA-2281913.  
XX PR 17-SEP-1999; 99CA-2281913.  
XX PA (KAYW/) KAY W W.  
XX PA (BURI/) BURIAN J.  
XX PA (KUZV/) KUZV M A.  
XX

PI Kay WM, Burian J, Kuzyk MA;  
XX  
DR WPI; 2001-316844/34.  
XX N-PSDB; AAF86248.  
XX  
PT Method for protecting polkilothermic fish against salmonid rickettsial  
PT septicaemia and other rickettsial diseases comprises administering a  
PT vaccine containing the Ospa protein of *Piscirickettsia salmonis* -  
XX  
PS Example 4; Fig 5; 35pp; English.  
XX  
CC This invention relates to a method for the protection against infection  
CC of a polkilothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC *salmonis*. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed Ospa, or an immunogenic fragment of a  
CC Ospa in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkilothermic fish, against the bacterial pathogen  
CC *P. salmonis*. The method is also useful for protecting against salmonid  
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
CC sequence represents the amino acid sequence of C17E2, a *P. salmonis* Ospa  
CC construct optimised for expression in *Escherichia coli*, fused to an  
CC undefined N-terminal fusion partner. The fusion protein is used in a  
CC vaccine to create an anti-Ospa antibody response.  
XX  
SQ Sequence 256 AA;  
  
Query Match 97.5%; Score 815; DB 22; Length 256;  
Best Local Similarity 98.8%; Pred. No. 9.3e-78;  
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
OY 3 RGCLQGSSLIITISVFLGCAQNFSEVGAATGAVGVAGOLFQKSGRYMAIGGAVL 62  
DB 97 RGCLQGSSLIITISVFLGCAQNFSEVGAATGAVGVAGOLFQKSGRYMAIGGAVL 156  
  
OY 63 GGLIGSKIGSMQDQDKIKLNQSLKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 122  
DB 157 GGLIGSKIGSMQDQDKIKLNQSLKVKAGVTRWRNPDTGNSYSVEPVRTYQRYNKOER 216  
  
OY 123 RQCYCREFOOKAMTAGOKOETIYGTACPODGRWIVISTEK 162  
DB 217 RQCYCREFOOKAMTAGOKOETIYGTACPODGRWIVISTEK 256  
  
RESULT 5  
AAB81130  
ID AAB81130 standard; Peptide: 20 AA.  
AC AAB81130;  
XX  
DT 11-JUL-2001 (first entry)  
XX  
DE Ospa B-cell epitope peptide #2.  
XX  
KW Polkilothermic fish; *Piscirickettsia salmonis*; rickettsial pathogen;  
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
KW SRS; antibody.  
XX  
OS *Piscirickettsia salmonis*.  
OS  
XX  
XX CA2281913-A1.  
XX PN  
XX PD 17-MAR-2001.  
XX PF 17-SEP-1999; 99CA-2281913.  
XX PR 17-SEP-1999; 99CA-2281913.  
XX PA (KAYW/) KAY W W.  
XX PA (BURI/) BURIAN J.  
XX PA (KUZV/) KUZV M A.  
XX PI Kay WM, Burian J, Kuzyk MA;  
XX

DR WPI: 2001-316844/34.

XX Method for protecting poikilothermic fish against salmonid rickettsial  
PT septicemia and other rickettsial diseases comprises administering a  
XX vaccine containing the OSPA protein of *Piscirickettsia salmonis*  
PS Example 2; Page 17; 35pp; English.

XX This invention relates to a method for the protection against infection  
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed OSPA, or an immunogenic fragment of a  
CC particularly poikilothermic fish, against the bacterial pathogen  
CC *P. salmonis*. The method is also useful for protecting against salmonid  
CC rickettsial septicemia (SR) and other rickettsial diseases. The present  
CC sequence represents an immunogenic epitope of the *P. salmonis* OSPA  
XX protein. The peptide is used to raise rabbit anti-OSPA antibodies.  
SO Sequence 20 AA;

Query Match

Best Local Similarity 13.4%; Score 112; DB 22; Length 20;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 110 PVRTQRYNKKQRRQOYCRE 129  
Db 1 PVRTQRYNKKQRRQOYCRE 20

RESULT 6

AAB20105

ID AAB20105 standard; Protein; 224 AA.

AC AAB20105;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB113 protein.

KW BASB113; infection; otitis media; pneumonia; therapy; diagnosis;

XX antibacterial; antimicrobial.

OS Moraxella catarrhalis.

PN WO200100836-A1.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000MO-EP05851.

XX 25-JUN-1999; 99GB-0015044.

PR (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PA Thonnard J;

XX WPI: 2001-112458/12.

DR N-PSDB; AAF30043.

XX

PT New BASB113 polypeptide isolated from Moraxella catarrhalis bacterium,

PT useful for diagnosing and producing vaccines against bacterial

XX infections such as otitis media and pneumonia

PS Claim 1; Page 67; 86pp; English.

XX The present sequence is that of BASB113 protein from Moraxella

CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis

CC media in children and pneumonia in adults. The invention provides

CC BASB113 polypeptides, and polynucleotides encoding them, as well as

CC expression vectors, host cells and methods for producing BASB113

CC polypeptides using recombinant methods. Also claimed is a vaccine

CC composition comprising a BASB113 polypeptide, an immunogenic

CC

CC Fragment of a BASB113 polypeptide, or a polypeptide having at least  
CC 85% amino acid sequence identity to BASB113, or comprising a  
CC polynucleotide encoding such a polypeptide. A claimed method of  
CC diagnosing a Moraxella infection involves identifying a BASB113  
CC polypeptide or antibody. A claimed therapeutic composition useful  
CC in treating humans with M. catarrhalis infection comprises at least  
CC 1 antibody directed against a BASB113 polypeptide. BASB113  
CC polypeptides also have utility in raising specific antibodies.  
XX and in screening for antibacterial drugs.

SO Sequence 224 AA;

Query Match

Best Local Similarity 12.9%; Score 108; DB 22; Length 224;  
Matches 36; Conservative 16; Mismatches 35; Indels 24; Gaps 5;

OY 8 GSSLIIIVELVGCANFRQEVGATG-----AVGGVAGOLFQSGSRVMAIG 59

Db 7 GYVLLASSMALAGCANTGT---TGNGTGGANVNRKAVIGAVAGAL---GCTATSKATG 60

OY 60 -----AVLGLISKTIGOSMDQDKIKLNOSLEKVKAGQYTRMRNPDTGN 104

Db 61 EKTGRDAILGAAGAAGAYMERQAK---QTEQMGQGTGVTHDPTGN 107

RESULT 7

AAY34487

ID AAY34487 standard; Protein; 223 AA.

AC AAY34487;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG3.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

XX vaccine; antigenic.

OS Porphyromonas gingivalis.

PN WO9929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98MO-AU01023.

XX 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

XX 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

XX 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

XX 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

XX

PA (CSLC-) CSL LTD.

XX

PI Agius CT, Barr IG, Hocking DM, Margolis MB, Patterson MA.

XX Ross BC, Rothel LJ, Webb EA.

XX

DR WPI: 1999-385613/32.

XX N-PSDB; AAX91705.

XX

PT Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

PS Claim 1; Page 469; 588pp; English.

XX AAY91536 to AAX91801 encode two hundred and sixty six antigenic

CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to

CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the



CC used as vaccines especially against *Porphyromonas gingivalis*. Probes can

Qy 30 VGATGAVVGVAGQLFEGKSGRVAMAIGGAVLGLIGSKIGQSDMOODKIKL 82

DB 62

VGTGSGALGAGTIGGSGGATAAVGAIGAAGAGTIEKMGVNGAEL 114

RESULT 10

ABG15906

ID ABG15906 standard; Protein; 309 AA.

AC ABG15906;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15897.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YF;

DR WPI: 2001-639362/73.

XX N-PSDB; AAS80093.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 20; SEQ ID No 46265; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence

CC tags (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on mutations

CC amino acid sequences. Ab000010-ABG30377 represent novel human

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 309 AA:

QY Query Match

DB Best Local Similarity 11.7%; Score 98; DB 22; Length 309;

Matches 33; Conservative 11; Mismatches 39; Indels 14; Gaps 3;

DB 13 IISVFLVGCANFRS-----DEVAATGAVGVAGVAGLFGKG--SGRVMAATGAVLGL 65

DB 185 IISIPAKVAVDNGKNKMAQAFGALIGAVAGGVIGHNVGSSNGTTRAGVAGGAVCAA 244

QY 66

IGSKI-----GOSMDQDKIKINOSLEKRVKAGQVT 95

DB 245 AGSMVNDKTLMEGSLTYKKEGTKYTSTGVCKEFOFT 281

RESULT 11

AAB82611

ID AAB82611 standard; Protein; 528 AA.

AC AAB82611;

DT 02-OCT-2001 (first entry)

DE Spider recombinant silk protein PETNCS.

KW Spider; orb-weaver; silk protein; PETNCS; structural protein;

XX purification; fibre; spinning.

XX Nephila clavipes.

XX Key Location/Qualifiers

FT Misc-difference 417 /note= "encoded by TAT"

FT Misc-difference 427 /note= "encoded by CGA"

FT Misc-difference 522 /note= "encoded by GAG"

XX WO20015333-A1.

PD 26-JUL-2001.

PF 01-NOV-2000; 2000WO-US30086.

XX 20-JAN-2000; 2000US-0490291.

PA (MELL/) MELLO C M.

PA (ARCT/) ARCIDIACONO S.

PA (BUTL/) BUTLER M M.

PA (USSA ) US SEC OF ARMY.

PI Mello CM, Arcidiacono S, Butler MM;

DR WPI: 2001-483136/52.

XX N-PSDB; AAB26304.

PT Recovering structural polypeptides in a biological sample, useful for  
PT purifying and spinning spider silks and other structural proteins,  
PT comprises treating the sample containing the polypeptides with an acid

XX Claim 2; Page 41-42; 49pp; English.

CC The present sequence is that of the orb-weaver spider (Nephila

CC clavipes) recombinant silk protein PETNCS. The invention

CC structural proteins. Organic acids are used to lyse recombinant

CC cells or other biological samples (such as non-recombinant

CC derived cells), and enrich the purity and yields of structural

CC proteins by hydrolysing many of the macromolecules while leaving

CC the structural proteins intact. In the case of silk proteins, the

CC resulting lysate is further purified by ion-exchange or affinity

CC chromatography and processed into an aqueous-based mixture for

CC into vector PET24 for recombinant expression in Escherichia coli,

CC containing denaturant (3 M guanidine-HCl) and by affinity

CC chromatography on nickel-NTA agarose. Products obtained using the

CC methods of the invention can be used in the construction of many

CC materials including films, fibres, woven articles, sutures, the new

CC ballistic protection, parachutes and parachute cords. The new

CC method has the following advantages over prior art: it involves

CC fewer steps, requires less time and smaller volumes of reagents.

CC results in better recovery of protein at higher purity (70-99%),  
CC is easy to scale up, and the fibres are spun in an environmentally  
CC benign solution reducing hazardous waste accumulation and cost.

SO Sequence 528 AA:

Query Match 10.9%; Score 91; DB 22; Length 528;

Best Local Similarity 51.1%; Pred. No. 0.51; Mismatches 19; Indels 2; Gaps 1;

Matches 24; Conservative 2; Mismatches 19; Indels 2; Gaps 1;

QY 28 QEVGATGAVGVAGQ--LFGKSGRVAMAGAVLGLIGSKIGQ 72

Db 320 QGAGAAAAAGGAGCGGCGAGAAAAAGAGCGGCGGAGCQ 366

RESULT 12

ABB66232 standard; Protein; 2309 AA.

XX ABB66232;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 25488.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX WPI: 2001-656860/75.

XX N-PSDB; ABLI0335.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 25488; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABLI01840-ABLI16175) and the encoded proteins

XX (ABBS7737-ABBS7072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 2309 AA:

Query Match 10.3%; Score 86.5; DB 22; Length 2309;

Best Local Similarity 27.4%; Pred. No. 9.9; Mismatches 55; Indels 17; Gaps 5;

Matches 34; Conservative 18; Mismatches 55; Indels 17; Gaps 5;

QY 23 QNFSROEVGAATGA---VVGVAAGQLFGKSGRVAMAGAVLGLIGSKIGSMDQDK 79

Db 2124 QQQQQRQVGGGNGPSMNLALGRSGAVGSGSGN-----GGGGGGGAGGSGVGGGNGV 2178

QY 80 IKLNQSLKXKAGQVTRWRNP-DTNGNSYSEPVRTYORYNKQERRQYCFEFOQKAMIAQ 138

Db 2179 GSVGQS---GGGQGRTPPIORPNYPQHP-----000000000000RE0AAAAVAA 2230

QY 139 QKOE 142

Db 2231 QRAQ 2234

RESULT 13

ABB58019 standard; Protein; 666 AA.

XX ABB58019;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 849.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX WPI: 2001-656860/75.

XX N-PSDB; ABLI02122.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 849; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABLI01840-ABLI16175) and the encoded proteins

XX (ABBS7737-ABBS7072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 666 AA:

Query Match 9.8%; Score 82; DB 22; Length 666;

Best Local Similarity 22.1%; Pred. No. 6.1; Mismatches 58; Indels 26; Gaps 2;

Matches 30; Conservative 22; Mismatches 58; Indels 26; Gaps 2;

QY 31 GAATGAVGVAGQLFGKSGRVAMAGAVLGLIGSKIGSMDQDKIKLNQSLKXK 90

Db 463 GPGTGGAGGGGSSSSNGNGTAAVS-GSVTTGAGAGSGTSSNSNSNGSGSGSVG 521

QY 91 AGQVTRWRNPDTGNSYSV-----EPVRYORYNKQERRQ 125

Db 522 SGSGSHPGTPTSLHAHSANGTSSSLGGLHLATPHQMVAAAGSPVVLHQ000000HQ00 581



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:09:08 : Search time 962.107 Seconds  
(without alignments)  
6775.784 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atcgctgctgcctgcagcaggg.....agggtgattagcaccgaaaaa 483

Scoring table: IDENTITY\_NUC

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
EST : \*  
1: em\_estdb : \*  
2: em\_esthum : \*  
3: em\_estin : \*  
4: em\_estu : \*  
5: em\_estov : \*  
6: em\_estpl : \*  
7: em\_estro : \*  
8: em\_hic : \*  
9: gb\_estl : \*  
10: gb\_estl2 : \*  
11: gb\_hic : \*  
12: gb\_gss : \*  
13: em\_gss\_hum : \*  
14: em\_gss\_huv : \*  
15: em\_gss\_pln : \*  
16: em\_gss\_vrt : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	42	8.7	238	10	BG240789	BG240789 OV1_38_CO
2	42	8.7	402	10	BG356895	BG356895 OV2_11_D0
3	42	8.7	419	10	BE355894	BE355894 DG1_11_E1
4	42	8.7	428	10	BF176742	BF176742 EM1_4_F01
5	42	8.7	446	10	BG052290	BG052290 RH12_12
6	42	8.7	500	10	BF586874	BF586874 FM1_31_F0
7	42	8.7	507	10	BG053591	BG053591 RH12_11
8	42	8.7	531	10	BG053014	BG053014 RH12_16
9	42	8.7	537	10	BF587805	BF587805 FM1_20_CO
10	42	8.7	561	10	BG673858	BG673858 EM1_20_CO
11	42	8.7	577	10	BE356763	BE356763 DG1_12_B1
12	42	8.7	582	10	BE356800	BE356800 DG1_12_E1
13	42	8.7	591	10	BE356800	BE356800 DG1_12_E1
14	42	8.7	597	10	BG102589	BG102589 RH12_34
15	42	8.7	598	9	AW672446	AW672446 LG1_359_D
16	42	8.7	600	10	BE360868	BE360868 DG1_67_A0
17	42	8.7	620	10	BF481524	BF481524 FM1_21_B0

18	42	8.7	649	10	BE355895	BE355895 DG1_11_E1
19	40.4	8.4	548	10	BM328366	BM328366 PIC1_28_D
20	40	8.3	513	9	AV430609	AV430609 AV430609
21	40	8.3	768	12	AQ690676	AQ690676 nbx00082B
22	38	7.9	910	12	CNS0060N	AL065629 Drosophila
23	38	7.9	1101	12	CNS0100Y	AL098380 Drosophila
24	37.8	7.8	595	10	BI607335	BI607335 RH74382.5
25	37.8	7.8	826	10	BI522679	BI522679 603175711
26	37.8	7.8	925	12	CNS0091P	AL053013 Drosophila
27	37.6	7.8	1344	10	BE364680	BE364680 601658216
28	37.4	7.7	350	10	BG052289	BG052289 RH12_12
29	37.4	7.7	1101	12	CNS017SY	AL108460 Drosophila
30	37.2	7.7	514	10	BG320949	BG320949 ZM04_02E0
31	37	7.7	908	10	BF253465	BF253465 HVSWEf000
32	36.8	7.6	446	9	AU182544	AU182544 AU182544
33	36.8	7.6	1101	9	AL513871	AL513871 AL513871
34	36.4	7.5	1043	10	BG104511	BG104511 602311153
35	36.2	7.5	595	12	BM478914	BM478914 BCGMR38TF
36	36.2	7.5	647	10	BJ003370	BJ003370 BJ003370
37	36.2	7.5	663	10	BJ003303	BJ003303 BJ003303
38	36.2	7.5	789	9	AU125061	AU125061 AU125061
39	36	7.5	1058	10	BG295311	BG295311 602392604
40	35.8	7.4	217	10	BM445199	BM445199 EBem09_S0
41	35.8	7.4	682	10	BE593319	BE593319 WS1_100_C
42	35.6	7.4	467	9	AU182484	AU182484 AU182484
43	35.6	7.4	586	10	BF978932	BF978932 602147607
44	35.6	7.4	653	10	BE293240	BE293240 601143811
45	35.6	7.4	685	9	AU130542	AU130542 AU130542

## ALIGNMENTS

RESULT 1  
LOCUS BG240789 238 bp mRNA linear EST 15-FEB-2001  
DEFINITION OV1\_38.C06.g1.A002 ovary 1 (OV1) Sorghum bicolor cDNA, mRNA  
SEQUENCE  
BG240789

ACCESSION BG240789.1 GI:12775862  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade: Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 238)

REFERENCE  
AUTHORS Cordonnier-Pratt/M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
L.H.

TITLE An EST database from Sorghum: ovaries of varying immature stages  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805

FEATURES  
source  
Seq primer: PolyTMix  
High quality sequence start: 3  
High quality sequence stop: 230  
POLYA-NO.  
Location/Qualifiers  
1..238  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/note="Organ: Mix of ovaries of varying immature stages  
from 8-week-old plants; Vector: pBluescript II from lambda  
zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made

BASE COUNT 63 a 54 c 70 g 51 t  
 Clones to be sequenced were prepared by mass excision."

Query Match  
 Best Local Similarity 59.0%; Pred. No. 0.19;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTTCAGCCGAGAGTTGGCGGCCACCGGTCGCTGTGGCGGCTGT 119  
 DB 79 CGCGGAGAACTTTCAGCCGAGAGTTGGCGGCCACCGGTCGCTGTGGCGGCTGT 119  
 QY 120 TGCCGCGCAGCTGTGGTAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 138  
 DB 139 CGACGCGCAGATCACTATGAAGAGTTGTAAAGTTATATGATGACCAAGTAGAGGAGCGGT 179  
 QY 180 TC 181  
 DB 199 CC 200

RESULT 2  
 BG356895  
 LOCUS 402 bp mRNA linear EST 06-MAR-2001  
 DEFINITION OY2.11.D01.g1.A002 Ovary 2 (OY2) Sorghum bicolor cDNA, mRNA  
 ACCESSION BG356895  
 VERSION BG356895.1 GI:13238881  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 1 (bases 1 to 402)  
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 L.H.  
 An EST database from Sorghum: ovaries of varying immature stages  
 unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu  
 Seq primer: PolyTMix  
 High quality sequence start: 47  
 High quality sequence stop: 391  
 POLYA-No.

TITLE An EST database from Sorghum: ovaries of varying immature stages  
 JOURNAL unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu  
 Seq primer: PolyTMix  
 High quality sequence start: 47  
 High quality sequence stop: 391  
 POLYA-No.

## FEATURES

source Location/Qualifiers

1..402  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /note="Organ: Ovary 2 (OY2)"  
 from 8-week-old plants; Vector: pluscript II from lambda  
 Zap II; Site:1: XhoI; Site:2: EcoRI; The library was made  
 from poly-A RNA in the cloning vector lambda Zap II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 107 a 82 c 117 g 96 t  
 ORIGIN

Query Match  
 Best Local Similarity 59.0%; Pred. No. 0.24;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTTCAGCCGAGAGTTGGCGGCCACCGGTCGCTGTGGCGGCTGT 119  
 DB 140 CGCGGAGAACTTTCAGCCGAGAGTTGGCGGCCACCGGTCGCTGTGGCGGCTGT 119  
 QY 120 TGCCGCGCAGCTGTGGTAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 138  
 DB 139 CGACGCGCAGATCACTATGAAGAGTTGTAAAGTTATATGATGACCAAGTAGAGGAGCGGT 179  
 QY 180 TC 181  
 DB 199 CC 200

DB 200 CGACGCGCAGATCACTATGAAGAGTTGTAAAGTTATATGATGACCAAGTAGAGGAGCGGT 259  
 QY 180 TC 181  
 DB 260 CC 261

RESULT 3  
 BE355894  
 LOCUS 419 bp mRNA linear EST 20-JUL-2000  
 DEFINITION DGI.11.E11.g1.A002 Dark grown 1 (DGI) Sorghum bicolor cDNA, mRNA  
 ACCESSION BE355894  
 VERSION BE355894.1 GI:9297451  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 1 (bases 1 to 419)  
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 L.H.  
 An EST database from Sorghum: dark-grown seedlings  
 unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu  
 Seq primer: PolyTMix  
 High quality sequence start: 16  
 High quality sequence stop: 336  
 POLYA-No.

TITLE An EST database from Sorghum: dark-grown seedlings  
 JOURNAL unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu  
 Seq primer: PolyTMix  
 High quality sequence start: 16  
 High quality sequence stop: 336  
 POLYA-No.

## FEATURES

source Location/Qualifiers

1..419  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 lambda Zap; Site:1: XhoI; Site:2: EcoRI; The library was  
 made from poly-A RNA in the cloning vector lambda Zap II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 102 a 82 c 116 g 119 t  
 ORIGIN

Query Match  
 Best Local Similarity 59.0%; Pred. No. 0.25;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTTCAGCCGAGAGTTGGCGGCCACCGGTCGCTGTGGCGGCTGT 119  
 DB 78 CGCGGAGAACTTTCAGCCGAGAGTTGGCGGCCACCGGTCGCTGTGGCGGCTGT 119  
 QY 120 TGCCGCGCAGCTGTGGTAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 137  
 DB 138 CGACGCGCAGATCACTATGAAGAGTTGTAAAGTTATATGATGACCAAGTAGAGGAGCGGT 179  
 QY 180 TC 181  
 DB 199 CC 199

RESULT 4  
 BE176742  
 LOCUS 428 bp mRNA linear EST 31-OCT-2000  
 DEFINITION EM1.4.F01.g1.A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA

sequence.  
 accession Bf176742  
 version Bf176742.1 GI:11064652  
 keywords EST.  
 source Sorghum.  
 organism Sorghum bicolor  
 Eukaryota: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; PACC clade: Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 428)  
 Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.  
 An EST database from Sorghum: developing embryos  
 Unpublished (2000)  
 contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@atuga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
 Seq primer: PolyTmix  
 High quality sequence start: 32  
 High quality sequence stop: 427  
 POLYA-No.  
 Location/Qualifiers  
 1..428  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_1lb="Embryo 1 (EM1)"  
 /note="Organ: Embryos germinated for 24 hr. Vector: Bluescript II from Lambda zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 103 a 80 c 119 g 126 t  
 ORIGIN  
 Query Match 8.7%; Score 42; DB 10; Length 428;  
 Best Local Similarity 59.0%; Pred. No. 0.25;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 60 CGCCAGACTTCAGCCGCGAGAGTTGGCGCGCCAGCGGCTGTGGCGGTGT 119  
 || ||||| || || ||||| || || || ||||| || || || |||||  
 Db 58 CGCGAGAGACTGACCCAGCAGAGAGTGCAGAGATGATCCGTGAGCGTGCATGG 117  
 QY 120 TGGCGGCGACCTGTTCGTTAAGGCTCTGCTGCTGCATGGCCATCGCGGCTGGCGGT 179  
 || ||||| || || ||||| || || ||||| || || ||||| || || |||||  
 Db 118 CGAGCGCCAGATCAACTATGAGAGTTGTTAAGGTTATGATGGCAAGTGAAGCGGT 177  
 QY 180 TC 181  
 Db 178 CC 179  
 RESULT 5  
 BG052290 446 bp mRNA linear EST 25-JAN-2001  
 LOCUS RH122.12.G04.g1\_A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA  
 DEFINITION sequence.  
 accession BG052290  
 version BG052290.1 GI:12506830  
 keywords EST.  
 source Sorghum prolinguam.  
 organism Sorghum prolinguam.  
 Eukaryota: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; PACC clade: Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 446)  
 Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt,L.H.

TITLE  
 JOURNAL  
 COMMENT  
 An EST database from Sorghum: Sorghum prolinguam rhizomes  
 Unpublished (2000)  
 contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@atuga.edu  
 Seq primer: PolyTmix  
 High quality sequence start: 11  
 High quality sequence stop: 440  
 POLYA-No.  
 Location/Qualifiers  
 1..446  
 /organism="Sorghum prolinguam"  
 /db\_xref="taxon:132711"  
 /clone\_1lb="Rhizome2 (RH122)"  
 /note="Organ: Rhizomes; Vector: Bluescript II from Lambda zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 101 a 81 c 124 g 140 t  
 ORIGIN  
 Query Match 8.7%; Score 42; DB 10; Length 446;  
 Best Local Similarity 59.0%; Pred. No. 0.25;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
 QY 60 CGCCAGACTTCAGCCGCGAGAGTTGGCGCGCCAGCGGCTGTGGCGGTGT 119  
 || ||||| || || ||||| || || || ||||| || || || |||||  
 Db 51 CGCGAGAGACTGACCCAGCAGAGAGTGCAGAGATGATCCGTGAGCTGACGTGATGG 110  
 QY 120 TGGCGGCGACCTGTTCGTTAAGGCTCTGCTGCTGCATGGCCATCGCGGCTGGCGGT 179  
 || ||||| || || ||||| || || ||||| || || ||||| || || |||||  
 Db 111 CGAGCGCCAGATCAACTATGAGAGTTGTTAAGGTTATGATGGCAAGTGAAGCGGT 170  
 QY 180 TC 181  
 Db 171 CC 172  
 RESULT 6  
 BF586874 500 bp mRNA linear EST 12-DEC-2000  
 LOCUS Fm1.31.E06.g1\_A003 Floral-Induced Meristem 1 (Fm1) Sorghum  
 DEFINITION prolinguam cDNA, mRNA sequence.  
 accession BF586874  
 version BF586874.1 GI:11679198  
 keywords EST.  
 source Sorghum prolinguam.  
 organism Sorghum prolinguam.  
 Eukaryota: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; PACC clade: Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 500)  
 Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt,L.H.  
 An EST database from Sorghum: floral-induced meristems  
 Unpublished (2000)  
 contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@atuga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
 Seq primer: PolyTmix  
 High quality sequence start: 8  
 High quality sequence stop: 495

FEATURES POLYA-No.  
Location/Qualifiers  
1. 500

BASE COUNT Clones to be sequenced were prepared by mass excision."  
ORIGIN  
1. 500  
/organism="Sorghum prolinguam"  
/db\_xref="taxon:132711"  
/clone\_lib="Rhizome2 (RH122)"  
/note="Organ: Floral-induced Meristem 1 (FM1)"  
pbluescript II from Lambda Zap II; Site:1: XhoI; Site:2:  
15 days with 16 hr darkness and 8 hr light (flowering is  
induced by short-day conditions); 16 days after being  
returned to the greenhouse under natural long days during  
late April/early May, meristems were harvested The  
library was made from poly-A RNA in the cloning vector  
lambda Zap II. Clones to be sequenced were prepared by  
mass excision."

Query Match 8.7%; Score 42; DB 10; Length 500;  
Best Local Similarity 59.0%; Pred. No. 0.27;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCCACCGGCTTGTGGCGGT 119  
DB 89 CGCGGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCCACCGGCTTGTGGCGGT 119  
QY 120 TCGCGGCGAGCTGTTCGTAAGGCTCTGTCGTGTGATGCGCATCGCGGTGGGT 148  
DB 149 CGACGGCCAGATCAACTATGAAGAGTTTGTAAAGCTTATGATGGCAAGTGAAGCGGT 179  
QY 180 TC 181  
DB 209 CC 210

RESULT 7 507 bp mRNA linear EST 25-JAN-2001  
BG053591  
LOCUS  
DEFINITION RH122\_11.C03.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA  
ACCESSION BG053591  
VERSION BG053591.1 GI:12509440  
KEYWORDS EST.  
SOURCE Sorghum prolinguam.  
ORGANISM Sorghum prolinguam.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
1 (bases 1 to 507)  
Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt  
L.H.  
An EST database from Sorghum: Sorghum prolinguam rhizomes  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@pratt.uga.edu  
Seq primer: T7  
High quality sequence start: 95  
High quality sequence stop: 507  
POLYA=Yes.

FEATURES  
source Location/Qualifiers  
1. 507

/organism="Sorghum prolinguam"  
/db\_xref="taxon:132711"  
/clone\_lib="Rhizome2 (RH122)"  
/note="Organ: Rhizomes; Vector: pbluescript II from Lambda  
Zap II; Site:1: XhoI; Site:2: EcoRI; The library was made  
from poly-A RNA in the cloning vector lambda Zap II."

BASE COUNT Clones to be sequenced were prepared by mass excision."  
ORIGIN  
109 a 92 c 142 g 164 t

Query Match 8.7%; Score 42; DB 10; Length 507;  
Best Local Similarity 59.0%; Pred. No. 0.27;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCCACCGGCTTGTGGCGGT 119  
DB 54 CGCGGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCCACCGGCTTGTGGCGGT 119  
QY 120 TCGCGGCGAGCTGTTCGTAAGGCTCTGTCGTGTGATGCGCATCGCGGTGGGT 113  
DB 114 CGACGGCCAGATCAACTATGAAGAGTTTGTAAAGCTTATGATGGCCAACTGAAGCGGT 173  
QY 180 TC 181  
DB 174 CC 175

RESULT 8 531 bp mRNA linear EST 25-JAN-2001  
BG053014  
LOCUS  
DEFINITION RH122\_16.D04.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA  
ACCESSION BG053014  
VERSION BG053014.1 GI:12508270  
KEYWORDS EST.  
SOURCE Sorghum prolinguam.  
ORGANISM Sorghum prolinguam.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
1 (bases 1 to 531)  
Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt  
L.H.  
An EST database from Sorghum: Sorghum prolinguam rhizomes  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@pratt.uga.edu  
Seq primer: PolyT/Mix  
High quality sequence start: 42  
High quality sequence stop: 527  
POLYA-No.

FEATURES  
source Location/Qualifiers  
1. 531

/organism="Sorghum prolinguam"  
/db\_xref="taxon:132711"  
/clone\_lib="Rhizome2 (RH122)"  
/note="Organ: Rhizomes; Vector: pbluescript II from Lambda  
Zap II; Site:1: XhoI; Site:2: EcoRI; The library was made  
from poly-A RNA in the cloning vector lambda Zap II."  
Clones to be sequenced were prepared by mass excision."  
BASE COUNT 122 a 101 c 152 g 154 t 2 others  
ORIGIN

Query Match 8.7%; Score 42; DB 10; Length 531;  
Best Local Similarity 59.0%; Pred. No. 0.27;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCCACCGGCTTGTGGCGGT 119  
DB 121 CGCGGAGAACTTCAGCCCGCCAGGAAGTTGGCGGCCACCGGCTTGTGGCGGT 119  
QY 120 TCGCGGCGAGCTGTTCGTAAGGCTCTGTCGTGTGATGCGCATCGCGGTGGGT 180  
DB 181 CGACGGCCAGATCAACTATGAAGAGTTTGTAAAGCTTATGATGGCCAACTGAAGCGGT 240



FEATURES	source	location/Qualifiers
BASE COUNT	118 a 98 c 152 g 159 t	
ORIGIN		
Query Match	8.7%: Score 42; DB 10; Length 537;	
Best Local Similarity	59.0%: Pred. NO. 0.27; 50; Indels 0; Gaps 0;	
Matches	72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;	
QY	60 CGCCGAGAACTTTCAGCCCGCAGGAAGTGGCGCGCCACCGGTCGGTGTGGCGGTGT	119
DB	92 CGCGGAGAAAGTCAGCGACGAGAGAGTGCAGAGATGATCCGTGAGGCTGAGCTGATGG	151
OY	120 TGCAGCGCAAGCTGTTGGTAAAGGCTGTGTCGTGTGTCGATGGCCATTCGGCGTGGCT	179
DB	152 CGACGCGCAAGATCACTATGAAGAGTTGTATAGCTTATGATGAGCCCAAGTGAAGAGCGGT	211
OY	180 TC 181	
DB	212 CC 213	
RESULT 10		
LOCUS	BF587805	561 bp mRNA linear EST 08-MAY-2000
DEFINITION	FM1_40_G05.g1_A003 Floral-induced Meristem 1 (FM1) Sorghum	
ACCESSION	BF587805	propiquinum cDNA, mRNA sequence.
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION		
LOCUS		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
DEFINITION	</	

VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	JOURNAL	COMMENT	REFERENCE	AUTHORS
BE587805.1	GI:11680115	EST	Sorghum propinquum.	An EST database from Sorghum: floral-induced meristems	Unpublished (2000)	Contact: Cordonnier-Pratt MM	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Department of Botany	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	The University of Georgia	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Tel: 706 542 1860	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Fax: 706 542 1805	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Email: mpratt@uga.edu	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Sequences have been trimmed to exclude PolyA, vector and regions	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	below Phred quality 16. The threshold for highest quality sequence	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	is 20.	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Seq primer: POLYTMix	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	High quality sequence start: 81	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	High quality sequence stop: 561	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	POLYA=NO.	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Location/Qualifiers	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	1..561	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	/organism="Sorghum propinquum"	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	/db_xref="taxon:132711"	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	/clone_lib="Floral-Induced Meristem 1 (FMI)"	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	/note="Organ: Floral-Induced meristems; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	BASE COUNT	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	140 a 111 c 160 g 150 t	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	ORIGIN	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Query Match	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Best Local Similarity 59.0%; Pred. No. 0.28;	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	60 CGCCAGAACTTCAGCCCGCAGAACTTGCGCGCCACCGGTGGCTGGCGGGGT 119	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	202 CGCGCAGAGCTGACCGCAGAGAGAGCTGACAGATGATCCGTGAGCTGACGTGATGG 261	1 (bases 1 to 561)	Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
BE587805.1	GI:11680115	EST	Sorghum propinquum.	Unpublished (2000)	Contact: Cordonnier-Pratt MM	120 TGCCGCGCAGCTGT		

REFERENCE 1 (bases 1 to 577)  
 AUTHORS Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.  
 TITLE An EST database from Sorghum: developing embryos  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu

Clade: Panicoideae; Andropogoneae; Sorghum.  
 (bases 1 to 577)  
 An EST database from Sorghum: developing embryos  
 Unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 57  
 High quality sequence stop: 572  
 POLYA=No.

#### FEATURES

source

Location/Qualifiers  
 1..577  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Embryo 1 (EM1)"  
 /note="Organ: Embryos germinated for 24 hr; Vector:  
 Bluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2:  
 EcoRI. The library was made from poly-A RNA in the cloning  
 vector Lambda Zap II. Clones to be sequenced were  
 prepared by mass excision."  
 BASE COUNT 129 a 106 c 166 g 176 t  
 ORIGIN

#### Query Match

Best Local Similarity 8.7%; Score 42; DB 10; Length 577;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGCAACTTCAGCCGCGAGAGTTGGCGCGCCCGCGGTGTGTGGCGGTGT 119  
 Db 121 CGCGGAGAGTGTGACGAGAGGAGGTCGACGATGATCCGTGAGGCTGACGTGATG 180  
 QY 120 TGCCGCGCAGCTGTGGTAAAGGCTGTGCTGTGATGAGGCGCATCGCGTGGCGT 179  
 Db 181 CGACGCGCAGATCACTATGAGAGTTGTAAAGTTATGATGCGCAAGTGAGAGCGGT 240  
 QY 180 TC 181  
 Db 241 CC 242

#### RESULT 12

BE356763

LOCUS

DEFINITION

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

BE356763 582 bp mRNA linear EST 20-JUL-2000  
 sequence.  
 BE356763  
 BE356763.1 GI:9298320  
 EST.  
 Sorghum.  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 Clade: Panicoideae; Andropogoneae; Sorghum.  
 Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt  
 L.H.  
 An EST database from Sorghum: dark-grown seedlings  
 Unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805

Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 6  
 High quality sequence stop: 553  
 POLYA=No.

#### FEATURES

source

Location/Qualifiers  
 1..582  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Dark Grown 1 (DG1)"  
 /note="Organ: 5-day-old dark-grown seedlings; Vector:  
 Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI. The library was  
 made from poly-A RNA in the cloning vector Lambda Zap II.  
 Clones to be sequenced were prepared by mass excision."  
 BASE COUNT 145 a 116 c 166 g 155 t  
 ORIGIN

#### Query Match

Best Local Similarity 8.7%; Score 42; DB 10; Length 582;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 60 CGCCGCAACTTCAGCCGCGAGAGTTGGCGCGCCCGCGGTGTGTGGCGGTGT 119  
 Db 213 CGCGGAGAGTGTGACGAGAGGAGGTCGACGATGATCCGTGAGGCTGACGTGATG 180  
 QY 120 TGCCGCGCAGCTGTGGTAAAGGCTGTGCTGTGATGAGGCGCATCGCGTGGCGT 179  
 Db 273 CGACGCGCAGATCACTATGAGAGTTGTAAAGTTATGATGCGCAAGTGAGAGCGGT 332  
 QY 180 TC 181  
 Db 333 CC 334

#### RESULT 13

BE356800

LOCUS

DEFINITION

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

BE356800 591 bp mRNA linear EST 20-JUL-2000  
 sequence.  
 BE356800  
 BE356800.1 GI:9298357  
 EST.  
 Sorghum.  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 Clade: Panicoideae; Andropogoneae; Sorghum.  
 Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt  
 L.H.  
 An EST database from Sorghum: dark-grown seedlings  
 Unpublished (2000)  
 Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: PolyTMix  
 High quality sequence start: 10  
 High quality sequence stop: 557  
 POLYA=No.

#### FEATURES

source

Location/Qualifiers  
 1..591  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 16:06:08 : Search time 35.3143 Seconds  
(without alignments)  
3359.570 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atgcgtgctgcctgcaggg.....agtgattagcaccgaaaaa 483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*
- 2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*
- 3: /cgn2\_6/prodata/2/lna/6A.COMB.seq:\*
- 4: /cgn2\_6/prodata/2/lna/6B.COMB.seq:\*
- 5: /cgn2\_6/prodata/2/lna/PCUTUS.COMB.seq:\*
- 6: /cgn2\_6/prodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.6	7.6	303	US-08-556-978B-80	Sequence 80, Appl
2	36.6	7.6	303	US-08-556-978B-81	Sequence 81, Appl
3	33.6	7.0	4403765	US-09-103-840A-2	Sequence 2, Appl
4	33.6	7.0	4411529	US-09-103-840A-1	Sequence 1, Appl
5	33.4	6.9	962	US-08-765-907A-16	Sequence 16, Appl
6	33.4	6.9	1052	US-08-403-852B-10	Sequence 10, Appl
7	33.4	6.9	1052	US-08-510-646B-10	Sequence 10, Appl
8	33.4	6.9	1052	US-09-231-818-10	Sequence 10, Appl
9	33.4	6.9	2888	US-08-765-907A-1	Sequence 1, Appl
10	33.2	6.9	23673	US-09-773-816-1	Sequence 1, Appl
11	33.2	6.9	4403765	US-09-103-840A-2	Sequence 2, Appl
12	33.2	6.9	4411529	US-09-103-840A-1	Sequence 1, Appl
13	33.2	6.8	1295	US-08-332-747-1	Sequence 1, Appl
14	32.4	6.7	535	5187077-4	Patent No. 5187077
15	32.4	6.7	535	5427925-3	Patent No. 5427925
16	32.4	6.7	599	5187077-24	Patent No. 5187077
17	32.4	6.7	599	5427925-22	Patent No. 5427925
18	32.4	6.7	714	5187077-16	Patent No. 5187077
19	32.4	6.7	1437	5427925-14	Patent No. 5427925
20	32.4	6.7	1437	5187077-16	Patent No. 5187077
21	32.4	6.7	1820	US-09-530-212A-1	Sequence 1, Appl
22	32.4	6.6	2365	US-08-363-208-1	Sequence 1, Appl
23	32.4	6.6	2365	US-09-137-478-1	Sequence 1, Appl
24	32.4	6.6	3300	US-08-194-377A-6	Sequence 6, Appl
25	32.4	6.6	3300	US-08-614-377A-6	Sequence 6, Appl
26	32.4	6.6	3300	US-09-142-648B-6	Sequence 6, Appl
27	31.8	6.6	13987	US-08-804-227C-13	Sequence 13, Appl

28	31.8	6.6	44377	US-08-804-227C-7	Sequence 7, Appl
29	31.8	6.6	44377	US-08-804-198-1	Sequence 1, Appl
30	31.6	6.5	5036	US-09-177-349-2	Sequence 2, Appl
31	31.2	6.5	7218	US-08-232-463-14	Sequence 14, Appl
32	30.8	6.4	1843	US-07-918-023-1	Sequence 1, Appl
33	30.8	6.4	2171	US-08-851-843A-100	Sequence 100, App
34	30.8	6.4	2171	US-08-974-549A-266	Sequence 266, App
35	30.8	6.4	2171	US-08-854-050-100	Sequence 100, App
36	30.8	6.4	2171	US-09-430-323-100	Sequence 100, App
37	30.6	6.3	882	US-08-818-112-138	Sequence 138, App
38	30.6	6.3	882	US-08-818-111-133	Sequence 138, App
39	30.6	6.3	882	US-09-056-556-138	Sequence 2, Appl
40	30.6	6.3	1689	US-08-442-884-2	Sequence 2, Appl
41	30.6	6.3	1820	US-08-173-508-7	Sequence 7, Appl
42	30.6	6.3	1821	US-08-265-310-7	Sequence 7, Appl
43	30.6	6.3	1821	US-08-951-742-7	Sequence 7, Appl
44	30.6	6.3	4030	US-09-293-505-1	Sequence 1, Appl
45	30.6	6.3	4705	US-07-998-003A-96	Sequence 96, Appl

## ALIGNMENTS

RESULT 1  
US-08-556-978B-80  
Sequence 80, Application US/08556978B  
Patent No. 6268169  
GENERAL INFORMATION:  
APPLICANT: FAHNESTOCK, STEPHEN F.  
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/556,978B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/077,600  
FILING DATE: JUNE 15, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: FLOYD, LINDA AXAMETHY  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CR-9389-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-8112  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-556-978B-80  
Query Match: 7.6%, Score 36.6, DB 4, Length 303,  
Best Local Similarity 57.4%, Pred. No. 0.043,  
Matches 66, Conservative 0, Mismatches 49, Indels 0, Gaps 0:  
QY GTTGGCGCGCCACCGCTGTTGGCGCGGTTCGCGCAGCTGTTGGAAGGC 144  
|||||



```

ADDRESSSEE: Flunegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

: ADDRESS: Finnegan, Henderson, Farabow, Garrett & Dunne
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
:

```





```
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Multisynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1
```

```
Query Match          6.9%; Score 33.4; DB 4; Length 2888;
Best Local Similarity 55.7%; Pred. No. 1.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
Oy 75 CCGCCAGAGATTGGCCGCCACCGGTGCGGTTGGCGGCTTCCGCGCAGCTGTT 134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2418 CGGCCAGCGAGTGTGTCAGGTCCCGCGGCGACCTGCGAGCGCTGGCGCCAGCAGCC 2359

Oy 135 CGGTAAGCGCTGCTGTGTCGATGCCATCGCGGCTGCGGTTGCGGCGGT 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2358 GGGTGTGCGCCGGCGGCGGTGCGGCTGATCCCGCGCGCTGCTGCCCGGT 2304
```

```
RESULT 10
US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; FILE REFERENCE: 28600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 23673
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23623)
; OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1
```

```
Query Match          6.9%; Score 33.2; DB 4; Length 23673;
Best Local Similarity 52.1%; Pred. No. 3.1;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```
Oy 55 GGTGGCCGCCAGAACTTCAGCCGCCAGAAATTGGCCGCCCGCGGCTGTGGGC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8366 GGGTTTCGGCAGATGACGTGCGACAGGATCGCAAGAGACGCCCGCGCGGTGAGT 8325

Oy 115 GGTGTTCCGCGCAGCTGTTGCGTAAGGCTCTGTGTGTCGATGGCCATCGGCGGT 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8326 GGGTTCGCTGACACGACGATGTGTAACACAGCCGCTCTGTCAGTTGCCAGAGGCGGC 8385

Oy 175 GCGGTTCTGGCGGCTGTGATG 196
    ||||| ||||| ||||| |||||
Db 8386 GCTGTCTTGGCCATCTGCATG 8407
```

```
RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
```

```
Query Match          6.9%; Score 33.2; DB 4; Length 4403765;
Best Local Similarity 50.6%; Pred. No. 23;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
Oy 59 GCGCCAGAACTTCAGCCGCCAGAACTTGCGCGGCCACCGGTGCGGCTGTG 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 835437 GCGCGCGGTCTCATTAACAGTCCGCTGCTCGACGATCAAGCGCCCTTCTGCGCGGT 835496

Oy 119 TTGCGCGCCAGCTGTGCGTAAGGCTCTGTGTCGATGCCATCGCGGCTGCGG 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 835497 TGGGCGCCCGCTGATGTGTAACGCGCCGACGAGGCGCGCGGAGCCCGCGCGGT 835556

Oy 179 TTCTGGCGGCTGTGATGCTTAATAATCGGTACAGC 216
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 835557 GGGCGCGCGATGTGTTGTCGCAACGCGCGCGCGGCG 835594
```

```
RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match          6.9%; Score 33.2; DB 4; Length 4411529;
Best Local Similarity 50.6%; Pred. No. 23;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
Oy 59 GCGCCAGAACTTCAGCCGCCAGAACTTGCGCGGCCACCGGTGCGGCTGTG 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

us-09-677-374-3.rni

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db  
422 CACAGACCCGCGGGTAAAGGCTGTGTC 151

QY 152 GTGTGTCATGGCCATCGCGGCTTCGTGGCGGTCGTGATTGGCTCTAAATCGGTC 211  
|| | ||| ||| ||| ||| | ||| | | ||| |||  
Db 362 TTGACCTGGAGGCTCAGCGCAGTGGGTTTCAGGACCTTCTGTCCCGGTGATATGCTC 303  
QY 212 AGAGCATGGACCAG 225  
|| | |||  
Db 302 AGGAGCGGCTCAG 289

Search completed: October 27, 2002, 20:15:39  
Job time : 5939.31 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 14:35:03 : Search time 122.627 Seconds  
(without alignments)  
6762.541 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483  
Sequence: 1 atgcgtgtgttcctgcaggg.....agtgattatgacgcgaaaaa 483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: N_Geneseq_032802.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	483	100.0	483	22	AAAF86247
2	483	100.0	768	22	AAAF86248
3	295	61.1	486	22	AAAF86246
4	295	61.1	489	22	AAAF86240
5	118	24.4	118	22	AAAF86254
6	108.4	22.4	110	22	AAAF86252
7	102	21.1	102	22	AAAF86255
8	92.4	19.1	94	22	AAAF86253
9	74	15.3	111	22	AAAF86251

#### ALIGNMENTS

Result	ID	Score	Query Match	Length DB	ID	Description
1	AAAF86247	483	100.0	483	22	AAAF86247
2	AAAF86247	483	100.0	768	22	AAAF86248
3	AAAF86246	295	61.1	486	22	AAAF86246
4	AAAF86240	295	61.1	489	22	AAAF86240
5	AAAF86254	118	24.4	118	22	AAAF86254
6	AAAF86252	108.4	22.4	110	22	AAAF86252
7	AAAF86255	102	21.1	102	22	AAAF86255
8	AAAF86253	92.4	19.1	94	22	AAAF86253
9	AAAF86251	74	15.3	111	22	AAAF86251

483; conservative 0;

483; conservative 0;

RESULT 3	
AAF68246	
ID	AAF68246 standard; DNA; 486 bp.
XX	
AC	AAF68246:
XX	
DT	11-JUL-2001 (first entry)
XX	
DE	Ospa 17kd antigen gene.
XX	
KW	Poljihothemic fish: Piscirickettsia salmonis; rickettsial pathogen;
KW	vaccine; Ospa, salmonid rickettsial septicemia; rickettsial disease;
KW	SRS; ds.
XX	
OS	Piscirickettsia salmonis.
XX	
FH	Key
FT	CDS
FT	1..486
FT	/+tag= a
FT	/partial
FT	/product= "Ospa"
FT	/note= "Genus specific 17kda antigen, the sequence does
FT	not include a stop codon"
FT	

WT: 2001-316844/34.  
P-PSDB: AAB81126.

**Method for protecting poikilothermic fish against salmonid rickettsial septicaemia and other rickettsial diseases** comprises administering a vaccine containing the Ospa protein of *Piscirickettsia salmonis*.

**Disclosure:** Fig 2B; 35pp; English.

This invention relates to a method for the protection against infection of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of a P. salmonis specific antigen termed Ospa, or an immunogenic fragment of Ospa in the form of a vaccine. The method is used for protecting animals, particularly poikilothermic fish, against the bacterial pathogen P. salmonis. The method is also useful for protecting against salmonid rickettsial septicaemia (SRS) and other rickettsial diseases. The present sequence represents DNA encoding the P. salmonis Ospa protein. An Ospa protein with an N-terminal fusion partner is used in a vaccine to create an anti-Ospa antibody response.

Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

	Query Match	61.1%; Best Local Similarity Matches 364; Conservative	Score 295; 76.0%; Pred. No. 2.9e-79; 0; Mismatches 115; Indels 0; Gaps 0;	DB 22; Length 486;
OY	5 GTGCTGGCCTCGAGGCAACGCTCTGCATCATTTATCTGTTCCTTGCGGGTTCGCCCC	64		
Db	8 GAGGATTGTTTGCMAAGTAGTACTGAATTAATTATAGTGTGTTTTAGTTGGCTGTGCC	67		
OY	65 AGAATTTCAGCGCGGAGAGAAGTTGGCGCGGCCACGCGGTGGGTGGCGGTGTCGCCG	124		
Db	68 AGAATTAGTGTGTAAGAAAGTCGAGAGCTGCGACATGGGGCTTTGTTGGCGGTGTTCG	127		
OY	125 GCCAGCTTTGCGTAAAGGCTCTGGCTGTGTGATGGCCATCGCGGTGGCTTCGG	184		
Db	128 GCCAGCTTTGTTGGTAAAGTAGTACTGTGCGATGCAATGGCCATTTGTGGTCTGTTTGG	187		
OY	185 GCGGTCTGATTTGGCTCTAAATCGTCAAGCATGATGGACCAGCAGATTAATCAACAATCA	244		
Db	188 GTGGATTTAATGGTCTCTAAATCGGTCTCAATCATCATGATCAGCAGATTAATAATTAAGCTPA	247		
OY	245 ACCACTCTCTGGAAAAGTGAAGAGCGCGGACAGTTACTGCTGGCGGTAAATCCGACACCG	304		
Db	248 ACCAGAGTTTGGAAAAGTGAAGAGCAGGSCAAGTACACAGCTTGGCGTAAATCCAGATACG	307		
OY	305 GTAACAGCTACTCTGTGSAACCGGTTTCGACCTTACAGAGCTTACAACAACAACAGAACGCC	364		
Db	308 GCATATGTTATATGTGTGAGCGACGTGCGTACTTACAGAGTTACAATTAAGCAAGAGCGTC	367		
OY	365 GTCAGCAGTACTGCCGCGAATTTTCAGCAGAAAGCATGATCGCAGGTCAAGAAACAGGAAA	424		
Db	368 GCCACCATTAATTTGTGCAAAATTTTCAGCAAAGGCGATGTACAGGCGCAAGAGAGA	427		
OY	425 TCTACGGACCGCGCGCCTCACGCCGAGTGGCGGTGAGAGGTGATTGACACCGAAAAA	483		
Db	428 TTTCAGGCACTGCATCCGCGCAACCGAGATGTCGTTGGCAAGTCAATTCACAGAAAAA	486		

RESULT 4  
AAH79040  
ID AAH79040 standard; cDNA; 489 BP.

XX AAH79040;  
XX  
XX  
XX 15-JAN-2002 (first entry)  
XX  
De Piscirickettsia salmonis polynucleotide P10.6.  
XX  
XX Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;  
XX septicaemia; SRS; surface antigen; vaccine; antibacterial; fish;  
XX ATCC VR-1361ss.

XX Piscirickettsia salmonis.  
XX

Key Location/Qualifiers  
1..489  
CDS /\*\*tag= a

WO20016865-A2.

20-SEP-2001.

12-MAR-2001; 2001WO-GB01055.

11-MAR-2000; 2000GB-0005838.

01-JUL-2000; 2000GB-0016080.

01-JUL-2000; 2000GB-0016082.

29-JUL-2000; 2000GB-0018599.

(AQUA-) AQUA HEALTH EURO LTD.

Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;  
Buzio L;

WPI: 2001-639050/73.

P-PSDB: AAG78025.

New nucleic acids encoding an amino acid sequence homologous to the  
surface antigen present on piscirickettsia salmonis are useful to  
protect fish against piscirickettsiosis.

Claim 3; Fig 4; 25pp; English.

The invention relates to nucleic acid sequences and the encoded protein  
of a least part of the surface antigen present on piscirickettsia  
salmonis for production of a vaccine with antibacterial activity to  
protect fish against P. salmonis which causes piscirickettsiosis, also  
known as salmonid rickettsial septicaemia.

Sequence 489 BP; 139 A; 79 C; 144 G; 127 T; 0 other;

Query Match 61.1%; Score 295; DB 22; Length 489;

Best Local Similarity 76.0%; Pred. No. 2.9e-79;

Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

5 GTGGTGGCTGCAGGCGCTCTGATCATTTATCTCTGTTGGTGGTGGCC 64

8 GAGATGTTTGCAGGATGATGATCTATTATATCATGTTGTTTATGCTGCC 67

AGACTTCACCGCCAGAGAGTTGGCGCGCCAGCGGTGTTGGCGGTTGCCG 124

AGACTTTAGTCTGCAAGAGTGGAGCTGGGACTGGGCTGTTGGCGGTTGCCG 127

125 GCCAGCTGTTGGTAAAGCTCTGGTCTGATGATGCCATCGCGGTGGTCTGG 184

128 GCCAGCTGTTGGTAAAGTGTGATGATGCAATGGCCATTTGGTGGTCTGG 187

185 GCGGCTGATGTTGCTAAATGCGTCAGAGCAGCAGCGATTAATCAACTGA 244

188 GTGATTAATGTTGTTAAATCGTCAATGATGATGATGATGATGATGATG 247

245 ACCAGTCTCTGAAAGAGTAAAGCCGCGAGTTACTGTTGGGTTATCGGACCG 304

248 ACCAGAGTTTGGAAAGTAAAGCAGAGGCAAGTGCAGCGTGGCGTATCAAT 307

305 GTACAGCTACTCTGTGGAACCGTTGCGACCTACACGTTTAAACAAACAGAC 367

308 GCAATAGTTTATGTTGTTGAGCCAGTCCGTTACTACACGTTTAAATTAAG 367

365 GTGAGCAGTACTGCGCGAATTTTCAGAGAAAGCAGATGATGCGAGTCAAG 424

368 GCCAGCAATATGTTGCGAGATTTTCAGCAAAAGCGATGATTTGCAAGGCA 427

425 TCTACGCAACCGCGTCCCTCAGCGGATGGCGCTGGCAGTGTGATTACAC 483

Db 428 TTTACGCACTGATGCCGCGACCGAGTGTGTTGGCAAGTCAATTTCACAGAAAAA 486

RESULT 5  
ID AAF86254 standard; DNA; 118 BP.

AAAF86254;

11-JUL-2001 (first entry)

PCR primer #4 used in cloning an optimisation of ospa gene.

Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;

vaccine; ospa; salmonid rickettsial septicaemia; rickettsial disease;

SRS; 17E2; fusion construct; PCR primer; ss.

Piscirickettsia salmonis.

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAYW/) KAY W W.

(BURJ/) BURIAN J.

(KUZJ/) KUZUK M A.

Kay WW, Burian J, Kuzuk MA;

WPI: 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial  
septicaemia and other rickettsial diseases comprises administering a  
vaccine containing the ospa protein of piscirickettsia salmonis

Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection  
of a poikilothermic fish by the bacterial pathogen, piscirickettsia  
salmonis. The method comprises administering an immunogenic amount of a  
P. salmonis specific antigen termed ospa, or an immunogenic fragment of a  
particularly poikilothermic fish, against the bacterial pathogen  
P. salmonis. The method is also useful for protecting animals,  
rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
sequence represents a PCR primer used in the cloning and optimisation of  
the P. salmonis ospa gene. The ospa gene is used in the method of the

Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;

Query Match 24.4%; Score 118; DB 22; Length 118;

Best Local Similarity 100.0%; Pred. No. 5.5e-26;

Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

208 GGTGAGCAGTGAACAGCAGATTAATCAACTGAACAGTCTGTGAAAGTGA 267

118 GGTGAGCAGTGAACAGCAGATTAATCAACTGAACAGTCTGTGAAAGTGA 59

268 GCGGCGAGTTATGCTGTTGGGTTATCGGACACCGTAAACAGTACTGTGGAAC 325

58 GCGGCGAGTTATGCTGTTGGGTTATCGGACACCGTAAACAGTACTGTGGAAC 1

RESULT 6  
ID AAF86252  
AAAF86252  
standard; DNA; 110 BP.



```

AC AAF86252;
XX
XX 11-JUL-2001 (first entry)
DE PCR primer #2 used in cloning an optimisation of Ospa gene.
XX
XX Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX PA (KAYW/) KAY W W.
XX PA (BURJ/) BURIAN J.
XX PA (KUZV/) KUZIK M A.
XX
XX PI Kay WW, Burian J, Kuzik MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
XX septicaemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
XX of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
XX salmonis. The method comprises administering an immunogenic amount of a
XX P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
XX Ospa in the form of a vaccine. The method is used for protecting animals,
XX particularly polkillothermic fish, against the bacterial pathogen
XX P. salmonis. The method is also useful for protecting against salmonid
XX rickettsial septicaemia (SRS) and other rickettsial diseases. The present
XX sequence represents a PCR primer used in the cloning and optimisation of
XX the P. salmonis Ospa gene. The Ospa gene is used in the method of the
XX invention.
XX
XX Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other:
SQ
Query Match 22.4%; Score 108.4; DB 22; Length 110;
Best Local Similarity 99.1%; Pred. No. 4.2e-23;
Matches 109; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 53 TGGGTTGGCCGACAACTTACAGCCGCGGAGAGTTGGCGGCGCCAGCGGTTGGG 112
DB 1 TGGGTTGGCCGACAACTTACAGCCGCGGAGAGTTGGCGGCGCCAGCGGTTGGG 60
OY 113 GCGGTTGGCCGCGGAGGCTTGGCTAAAGGCTTGGCTGCTGTCATG 162
DB 61 GCGGTTGGCCGCGGAGGCTTGGCTAAAGGCTTGGCTGCTGTCATG 110

```

## RESULT 7

AAF86255/c  
ID AAF86255 standard; DNA: 102 BP.

AC AAF86255;  
XX

DT 11-JUL-2001 (first entry)

DE PCR primer #5 used in cloning an optimisation of Ospa gene.

KM Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
KW SRS; 17E2; fusion construct; PCR primer; ss.

```

KM SRS; 17E2; fusion construct; PCR primer; ss.
XX
XX Piscirickettsia salmonis.
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX PA (KAYW/) KAY W W.
XX PA (BURJ/) BURIAN J.
XX PA (KUZV/) KUZIK M A.
XX
XX PI Kay WW, Burian J, Kuzik MA;
XX
XX WPI; 2001-316844/34.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
XX septicaemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis -
XX
XX Example 3; Fig 4B; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
XX of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
XX salmonis. The method comprises administering an immunogenic amount of a
XX P. salmonis specific antigen termed Ospa, or an immunogenic fragment of
XX Ospa in the form of a vaccine. The method is used for protecting animals,
XX particularly polkillothermic fish, against the bacterial pathogen
XX P. salmonis. The method is also useful for protecting against salmonid
XX rickettsial septicaemia (SRS) and other rickettsial diseases. The present
XX sequence represents a PCR primer used in the cloning and optimisation of
XX the P. salmonis Ospa gene. The Ospa gene is used in the method of the
XX invention.
XX
XX Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other:
SQ
Query Match 21.1%; Score 102; DB 22; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.6e-21;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 334 ACCTACGAGCGTTACACAAACAGAGAGCGCCGTCAGAGTACGCGCGAATTTCAGCAG 393
DB 102 ACCTACGAGCGTTACACAAACAGAGAGCGCCGTCAGAGTACGCGCGAATTTCAGCAG 43
OY 394 AAGGCGATGATCCAGGTCAGAAACAGAGAAATCTACGGCACC 435
DB 42 AAGGCGATGATCCAGGTCAGAAACAGAGAAATCTACGGCACC 1

```

RESULT 8  
AAF86253  
ID AAF86253 standard; DNA: 94 BP.

AC AAF86253;  
XX

DT 11-JUL-2001 (first entry)

DE PCR primer #3 used in cloning an optimisation of Ospa gene.

KM Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
KW SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

```

PF 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
PA (KAYW/) KAY W W.
PA (BURI/) BURIAN J.
PA (KUZV/) KUZYK M A.
PI Kay WW, Burian J, Kuzyk MA;
XX WPI; 2001-316844/34.
XX
PT Method for protecting poikilothermic fish against salmonid rickettsial
PT septicemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX
XX Example 3; Fig 4B; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
XX invention.
SQ Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;
Query Match 19.1%; Score 92.4; DB 22; Length 94;
Best Local Similarity 98.9%; Pred. No. 2.7e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 139 AAGGCTGTGCTGTGTCGATGCGCATGCGGCGGCTTGTGGCGGTGTGATTGGC 198
DB 1 AAGGCTGTGCTGTGTCGATGCGCATGCGGCGGCTTGTGGCGGTGTGATTGGC 198
QY 199 TCTAATATCGGTGACGATGACGACGAGATA 232
DB 61 TCTAATATCGGTGACGATGACGACGAGATA 94
RESULT 9
AAF86251
ID AAF86251 standard; DNA; 111 BP.
XX
AC AAF86251;
XX
DE 11-JUL-2001 (first entry)
XX
XX PCR primer #1 used in cloning an optimisation of Ospa gene.
XX
XX Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer; ss.
XX Piscirickettsia salmonis.
OS
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX PA (BURI/) BURIAN J.
XX PA (KUZV/) KUZYK M A.
XX

```

```

PI Kay WW, Burian J, Kuzyk MA;
XX WPI; 2001-316844/34.
XX
XX Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis
XX
XX Example 3; Fig 4B; 35pp; English.
XX
CC This invention relates to a method for the protection against infection
CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
CC Ospa in the form of a vaccine. The method is used for protecting animals,
CC particularly poikilothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the
XX invention.
SQ Sequence 111 BP; 16 A; 32 C; 30 G; 33 T; 0 other;
Query Match 15.3%; Score 74; DB 22; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCGTGTGCTGCTGCGAGGCGAGCTCTCTGATCATTTATCTCTGTTCTGGGGTTC 60
DB 38 ATGCGTGTGCTGCTGCGAGGCGAGCTCTCTGATCATTTATCTCTGTTCTGGGGTTC 60
QY 61 GCCCAGAACTTCAG 74
DB 98 GCCCAGAACTTCAG 111
RESULT 10
AAF86256/C
ID AAF86256 standard; DNA; 110 BP.
XX
AC AAF86256;
XX
DE 11-JUL-2001 (first entry)
XX
XX PCR primer #6 used in cloning an optimisation of Ospa gene.
XX
XX Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
XX SRS; 17E2; fusion construct; PCR primer; ss.
XX Piscirickettsia salmonis.
OS
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX PA (BURI/) BURIAN J.
XX PA (KUZV/) KUZYK M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
XX WPI; 2001-316844/34.
XX
XX Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the Ospa protein of Piscirickettsia salmonis
XX

```

PS Example 3; Fig 4b; 35pp; English.  
XX  
CC This invention relates to a method for the protection against infection  
CC of a polkiothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of  
CC OspA in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkiothermic fish, against the bacterial pathogen  
CC *P. salmonis*. The method is also useful for protecting against salmonid  
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
CC sequence represents a PCR primer used in the cloning and optimisation of  
CC the *P. salmonis* OspA gene. The OspA gene is used in the method of the  
CC invention.  
XX  
SO Sequence 110 BP; 17 A; 34 C; 27 G; 32 T; 0 other;  
  
Query Match 15.0%; Score 72.4; DB 22; Length 110;  
Best Local Similarity: 98.6%; Pred. No. 3.3e-12;  
Matches 73; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 410 GTCAGAAACGGAATCTACGCGACCGCGCTCCTCAGCCGCGATGCGCGCAGGTGA 469  
Db 110 GTCAGAAACGGAATCTACGCGACCGCGCTCCTCAGCCGCGATGCGCGCAGGTGA 51  
Qy 470 TTACACCGGAAAA 483  
Db 50 TTACACCGGAAAA 37  
  
RESULT 11  
AAC45883/c  
ID AAC45883 standard; DNA: 2292 BP.  
XX  
AC AAC45883;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 48114.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142970.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.

Query Match	Best Local Similarity	8.28;	Score 39.4;	DB 21;	Length 2292;
Matches	67; Conservative	0;	Mismatches 46;	Indels 0;	Gaps 0
QY	85 GTTGCCCGGCCACCCGTCGTTGTGGCGCGTGTGCCGSCAGCTGTGGTAAAGGC	144			
Db	1256 GTTGAGAGCTCTCCGCGTGTCTTTTGAAGAGTCTCTGTGTGTGTGGAACTGTC	1197			
QY	145 TCTGTCGTCGTGTGCATGTCGATCCGCGTCGCGTTCGTGAGCGCTGATATGG	197			
Db	1196 TCTGTGTGTCTTTTGGAGAGATTCCGCGTGTGTGTGTGGAGTCTTCTCG	1144			

AC  
AT18913;  
XX  
XX  
17-JAN-1997 (first entry)  
DDF  
DE  
DNA encoding spider dragline variant, DP-1B.9 monomer.  
XX  
XX  
Spider dragline

N	WO9429450-A2.
X	
D	22-DEC-1994.
D	
X	
F	15-JUN-1994;
X	94WO-US06689.
X	
3	15-JUN-1993;
3	93US-0077600.
3	

Fahnestock SR;  
WPI; 1995-036479/05  
P-PSDB; AAR99054.

CC This sequence encodes a synthetic spider dragline variant monomer,  
CC DP-1B.9. The sequence of the DP-1B.9 polymer is given in AAR99055.  
CC The polypeptide monomer is a variant based on a consensus sequence  
CC derived from the fibre forming regions of spider dragline protein,  
CC esp. the natural protein 1 (Spidroin 1) from *Nephila clavipes*.  
CC DNA sequence may be used in the recombinant production of the  
CC variant protein in a recombinant host, e.g. *E. coli* or *Bacillus*  
CC subtilis. Synthetic analogues of DP1 were designed to mimic the  
CC repeating consensus sequence of the natural protein and the pattern

CC	by reordering the four repeats within the monomer of DP-1A. This addition,
CC	monomer exhibits all of the regularities of (1)-(3) below. In addition,
CC	it exhibits a regularity of the natural sequence which is not shared by
CC	DP-1A, namely that a repeat in which both GYG and GRG are deleted is
CC	generally preceded by a repeat lacking the entire poly-alanine repeat,
CC	with one intervening repeat. The sequence of DP-1B matches the natural
CC	sequence more closely over a more extended segment than does DP-1A. The
CC	individual repeats differ from the consensus according to the pattern:
CC	(1) the poly-alanine sequence varies in length from 0-7
CC	residues; (2) when the entire poly-alanine sequence is deleted,
CC	so also is the surrounding sequence encompassing AGRGLGGGAGANG;
CC	(3) aside from the poly-alanine sequence, deletions usually
CC	encompass integral multiples of three consecutive residues;
CC	(4) deletion of GTC is generally accompanied by deletion of GRG
CC	in the same sequence; and
CC	(5) a repeat in which the entire poly-alanine sequence is
CC	deleted is generally preceded by a repeat containing six alanine
CC	residues.
CC	The proteins may be used to produce fibres of high tensile strength and
CC	elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC	for implants, plastic reinforcements, films, coatings, etc.
CC	
XX	
S0	Sequence 303 BP; 43 A; 69 C; 138 G; 53 T; 0 other:
QY	Query Match            7.6%; Score 36.6; DB 16; Length 303;
	Best Local Similarity   57.4%; Pred. No. 0.34;
Db	Matches   66; Conservative   0; Mismatches   49; Indels   0; Gaps   0;
QY	85 GTTGGCGGCGCACCGGTGGGTGTGGCGGTGTGGCGGCACGCTTTGCGTAAGGC 144
Db	87 GCTGGCGAGCGTGCTGT 146
QY	145 TCTGCTCCTGTGTGCATGCGCATGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 199
Db	147 GCTGTGCAAGCAGCAGCGTGCCTGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 201
RESULT 13	
AAT18912	
ID	AAT18912 standard; DNA; 303 BP.
XX	
AC	AAT18912:
XX	
DT	15-JAN-1997 (first entry)
XX	
DE	DNA encoding spider dragline variant, DP-1A.9 monomer.
XX	
KW	Spider; dragline protein; variant; monomer; polymer;
KM	fibre forming region; Spidroin I; Nephila clavipes; DPl; mimlc;
KW	DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KM	rope; surgical suture; implant; reinforcement; film; coating; ss.
XX	
CS	Synthetic.
XX	
PN	WO9429450-A2.
XX	
PD	22-DEC-1994.
XX	
PF	15-JUN-1994; 94WO-US06689.
XX	
PR	15-JUN-1993; 93US-0077600.
XX	
PA	(DUPO ) DU PONT DE NEMOURS & CO E.I.
XX	
PI	Fahnestock SR.
XX	
DR	WPI: 1995-036479/05.
XX	
DR	P-PSDB; AAR93052.
XX	
PT	New synthetic variants of spider dragline protein - for making
XX	fibres useful as clothing, surgical silk, plastic reinforcement
PT	etc., also related DNA, vectors and transformed cells

	xx	Claim 2; Page 121: 168bp; English.
	ps	
	xx	This sequence encodes a synthetic spider dragline variant monomer,
	CC	Dp-1A. <sup>9</sup> . The sequence of the Dp-1A. <sup>9</sup> polymer is given in AAR99053.
	CC	The polypeptide monomer is a variant based on a consensus sequence
	CC	derived from the fibre forming regions of spider dragline protein,
	CC	esp. the natural protein I (spidroin I) from Nephila clavipes. This
	CC	DNA sequence may be used in the recombinant production of the
	CC	variant protein in a recombinant host, <sup>e.g.</sup> E. coli or Bacillus
	CC	sutils. Synthetic analogues of DPI were designed to mimic the
	CC	repeating consensus sequence of the natural protein and the pattern
	CC	of variation among individual repeats. Dp-1A analogues are composed
	CC	of a tandemly repeated 101 amino acid monomer which comprises four
	CC	repeats which differ from the consensus sequence given in AAW06201,
	CC	according to the pattern (1)-(5) <sup>*</sup> given below. This 101 amino acid
	CC	monomer is repeated 1-16 times in a series of analogue proteins. The
	CC	individual repeats differ from the consensus according to the pattern:
	CC	(1) the poly-alanine sequence varies in length from 0-7
	CC	residues; (2) when the entire poly-alanine sequence is deleted,
	CC	so also is the surrounding sequence encompassing AKGGTGSQGAAGG;
	CC	(3) aside from the poly-alanine sequence, deletions usually
	CC	encompass integral multiples of three consecutive residues;
	CC	(4) deletion of GYG is generally accompanied by deletion of GRK
	CC	in the same sequence; and
	CC	(5) a repeat in which the entire poly-alanine sequence is
	CC	deleted is generally preceded by a repeat containing six alanine
	CC	residues.
	CC	The proteins may be used to produce fibres of high tensile strength and
	CC	elasticity, suitable for clothing, rope, surgical sutures, biomaterials
	CC	for implants, plastic reinforcements, films, coatings, etc.
	XX	
SQ	Sequence 303 BP; 43 A; 69 C; 138 G; 53 T; 0 other;	
Query Match	7.6%; Score 36.6; DB 16; Length 303;	
Best Local Similarity	57.4% ; Pred No. 0.34;	
Matches    66; Conservative     0; Mismatches    49; Indels        0; Gaps              0.		
Oy	85 GTTGCGCACCACGCCGTGCCTGTTGGCCGGTTCGCCGACCTGYTCGTAAGGC 144   Db 138 GTGGCGCACGTGGTAGCTGCTCAAGCGCGGTCTCATCAAGAAGGCCGGTCAAGSC 197  Oy 145 TCTGGTCGTGTGCATGGCCACTGCGCGCTTCCTGCGCGTCTATTGGCT 199   Db 198 GCTGGTGACGACGAGCTGCCCTGGCGGTGCAGGCCAAAGTGATATGGTGGCT 252	
RESULT 14		
AATJ31799/C		
ID AATJ31799 standard; DNA; 1305 BP.		
XX AC AATJ31799;		
DT 20-SEP-1996 (first entry)		
DE Erythrobaacter longus lycopen cyclase gene.		
KM Phycoene dehydrogenase: Erythrobaacter longus; Erwina heblcola; probe: KW open reading frame; lycopene cycloase; Rhodobaacter sphaeroides; KM beta-carotene; food colourant; additive; anti-cancer; ds.		
OS Erythrobaacter longus.		
PN JP08089241-A.		
PD 09-APR-1996.		
PF 30-SEP-1994; 94JP-0236621.		
PR 30-SEP-1994; 94JP-0236621.		
PA (ASAH ) ASAMI KASEI KOYO KK.		

DR WPI: 1996-23337/24.  
DR P-PSDB; AAR95698.  
XX Erythrobacter sp. phytoene dehydrogenase and lycopene cyclase genes  
PT - for the produ. of beta-carotene useful as a food colourant  
XX  
XX Example 1; Page 6-7; 7pp; Japanese.  
XX  
CC This is the nucleotide sequence encoding the lycopene cyclase enzyme  
CC from Erythrobacter longus ATCC 14126. The gene was isolated from an  
CC E. longus genomic DNA library using a 1 kb fragment of the phytoene  
CC dehydrogenase gene from Erwinia hebelicola as a probe. The probe isolated  
CC a nucleotide fragment which contained two open reading frames (ORF).  
CC ORF2 (1578 bp) encodes the phytoene dehydrogenase (AAT31798) whereas ORF1  
CC (1302 bp) encodes lycopene cyclase. The two ORFs were inserted into  
CC plasmid pRK415 to generate pRKCD. This plasmid was transformed into  
CC Rhodospirillum rubrum for production of the proteins. The phytoene  
CC dehydrogenase and lycopene cyclase can be used to produce beta-carotene  
CC for use as a food colourant or as a food additive with anti-cancer  
CC effects.  
XX  
XX Sequence 1305 BP; 251 A; 392 C; 403 G; 259 T; 0 other;  
SQ  
XX  
XX Query Match  
XX Best Local Similarity 7.5%; Score 36; DB 17; Length 1305;  
XX Matches 66; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 115 GTGTTGCCGCGCAGCTGTTGCGTAAGGCTGCGTGTGATGCGATCGCGCGT 174  
Db 900 GGTTAGTGTGTCAGCGGATGAGTAAACCGCGCGCGCGCGCGCGCGCGCGCG 841  
QY 175 GCGGTTCTGGGGGCTGCTGATTTGCTTAATTCGTCAGAGCATGACGACGAGA 230  
Db 840 AATGCGCACTTCTGCTGATAGCGGAAATTCGCCGCCGCGCGCGCGCGCGAGA 785  
RESULT 15  
ABLO3846  
ID ABL03846 standard; cDNA; 3251 BP.  
XX  
AC ABL03846;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6020.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR P-PSDB; ABB59743.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
PS Claim 1; SEQ ID NO 6020; 21pp + Sequence Listing; English.  
XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB57737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX Sequence 3251 BP; 859 A; 752 C; 758 G; 882 T; 0 other;  
SQ  
XX  
XX Query Match  
XX Best Local Similarity 7.5%; Score 36; DB 23; Length 3251;  
XX Matches 117; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
QY 52 GTGGTTGCCGCCAGAACTTCAGCGCGCGAGAGTTGGCGCGCGCGCGCGGTTG 111  
Db 1279 GTGATGATGTCACAGCAACTCGGGCTTCAGAGAGCGCCCTCGACAGAGCCATCTAT 1338  
QY 112 GCGGTTTGGCGCGCGCAGCTGTTGCGTAAGGCTGTCGTGTCGATGCGCATCGGC 171  
Db 1339 GTGCGGCTTCTTGGCGCAGCTCTTGGCGGCGGCGGCGGCGGCGGCGGCGGAT 1398  
QY 172 GTGCGGTTCTGCGCGGCTGATTGCTCTAAATCGGTCAGAGCATGAGCAGCAGAT 231  
Db 1399 GCGGAGGCTGCGGAGACACCTCTTGGAGATGTTGTCTGACGCACTGCGGAGGAGGC 1458  
QY 232 AAATCAACTGTAACCAAGCTCTGGAAGAAAGTGAAGCCGCGCAGTTACTGTTGGCT 291  
Db 1459 GTGACCTCTTGAAGCCCTGCAATTTGGAGGAAAGAAATATATCATTTGGAAGATTAGCT 1518  
QY 292 AATCGGACACC 303  
Db 1519 CTGCGGCGGCGACC 1530

Search completed: October 27, 2002, 16:40:46  
Job time : 129.627 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:07:23 : Search time 935.413 Seconds  
(without alignments)  
10805.408 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atgcgtgctgctgcaggg.....agtgattagcaccgaaaaa 483

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:\*

1: gb\_ba:\*

2: gb\_hlg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_svs:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: gb\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_ot:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_svs:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hlg\_hum:\*

31: em\_hlg\_inv:\*

32: em\_hlg\_other:\*

33: em\_hgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result Query Match Length DB ID Description

1	295	61.1	489	6	AX252413	AX252413 Sequence
2	295	61.1	493	1	AF184152	AF184152 Piscirickettsia
3	46.6	9.6	537	1	R1R17KCA	R1R17KCA
4	45	9.3	237523	1	RFX04	RFX04
5	42.8	8.9	203050	1	AL646071	AL646071 Ralstonia
6	41.6	8.6	448	1	RSU76907	RSU76907 Rickettsia
7	41	8.5	198050	1	AL646061	AL646061 Ralstonia
8	40.8	8.4	213050	1	AL646079	AL646079 Ralstonia
9	40.2	8.3	539	1	R1RANT17KA	R1RANT17KA
10	40.2	8.3	539	1	R1RANT17KD	R1RANT17KD
11	40.2	8.3	10127	1	AE008675	AE008675 Rickettsia
12	39.6	8.2	13946	1	AE004537	AE004537 Pseudomonas
13	39.4	8.2	106702	8	ATFXK2	ATFXK2
14	39.4	8.2	199577	8	ATCHRIV57	ATCHRIV57
15	39	8.1	1070	3	DM60844	DM60844
16	39	8.1	1070	3	DM60845	DM60845
17	39	8.1	1070	3	DM60847	DM60847
18	39	8.1	1070	3	DM60850	DM60850
19	38.6	8.0	535	1	R1RANT17KB	R1RANT17KB
20	38.6	8.0	620	1	R1RANT17K	R1RANT17K
21	38.4	8.0	125020	9	AF429315	AF429315 Homo sapi
22	38	7.9	2820	1	SV117269	SV117269 Streptomyces
23	37.8	7.8	13123	1	AE003872	AE003872 Xylella f
24	37.4	7.7	804	3	DSU60862	DSU60862 Drosophila
25	37.4	7.7	804	3	DSU60864	DSU60864 Drosophila
26	37.4	7.7	804	3	DSU60866	DSU60866 Drosophila
27	37.4	7.7	804	3	DSU60868	DSU60868 Drosophila
28	37.4	7.7	804	3	DSU60869	DSU60869 Drosophila
29	37.4	7.7	1070	3	DM60846	DM60846 Drosophila
30	37.4	7.7	1070	3	DM60848	DM60848 Drosophila
31	37.4	7.7	1140	1	AB046994	AB046994 Streptomyces
32	37.4	7.7	10029	1	AE008032	AE008032 Agrobacterium
33	37.4	7.7	10606	1	AE009066	AE009066 Agrobacterium
34	37.4	7.7	179855	2	AC106349	AC106349 Rattus norvegicus
35	37.2	7.7	416	1	AF031534	AF031534 Rickettsia
36	37	7.7	9811	14	AF449714	AF449714 Caulobacter
37	37	7.7	13348	1	AE005945	AE005945 Caulobacter
38	36.8	7.6	492	1	R1R1RAPRO	R1R1RAPRO M74042 R. australis
39	36.8	7.6	1073	3	DSU60861	DSU60861 Drosophila
40	36.8	7.6	6526	1	SCBAC14E8	SCBAC14E8 Streptomyces
41	36.8	7.6	56344	2	AC100687	AC100687 Mus musculus
42	36.8	7.6	138881	2	AP004589	AP004589 Oryza sativa
43	36.8	7.6	164632	2	AC025624	AC025624 Homo sapiens
44	36.8	7.6	169582	2	AC092934	AC092934 Homo sapiens
45	36.8	7.6	171052	2	AC025802	AC025802 Homo sapiens

#### ALIGNMENTS

RESULT 1	AX252413	489 bp	DNA	linear	PAT 05-OCT-2001
LOCUS	AX252413				
DEFINITION	Sequence 5 from Patent WO016865.				
ACCESSION	AX252413				
VERSION	AX252413.1	GI:15985721			
KEYWORDS					
SOURCE	Piscirickettsia salmonis.				
ORGANISM	Piscirickettsia salmonis				
REFERENCE	1 (bases 1 to 489)				
AUTHORS	Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and Burzio,L.				
TITLE	Fish vaccine against piscirickettsia salmonis				
JOURNAL	Patent: WO 016865-A 5 20-SEP-2001;				
FEATURES	Aqua Health (Europe) Limited (GB)				
SOURCE	Location/Qualifiers				
BASE COUNT	139 a 79 c 144 g 127 t				
ORIGIN					

Query Match	Best Local Similarity	61.1%	Score 295	DB 6	Length 489
Query	76.0%	Pred. No. 1	9e-61		
Matches 364	Conservative	0	Mismatches 115	Indels	Gaps
0					
5	GTGGTTGCCCTGACGGGACAGCTCTCTGATCAATATATCTGTCTTTCTGTCGGTGGTGGCCG	64			
8	GAGATGTTTGGCAAGGAGTAGTCTAATATATATTCAGTGTGTTTTTATGTTGGCTGTGGCC	67			
65	AGAACTTCACGCCCGACGAAGTTGGCGCGCCACCGGTGGGTGGGCGGTGTTCCG	124			
68	AGAACTTATGCTCAAGAAAGTCGGAGCTGCCAGCTGGGGCTGTCTTGGCGGTGTTCTG	127			
125	GCCAGCTGTTCCGTAAGGCTCTGTCGTCGTCTGTCATATGGCCATCGCGCGGTCTCTG	184			
128	GCCAGCTGTTTGGTAAAGGAGTACTGAGTTGCATATGGCCATATGTTGGTCTGTTTGG	187			
185	GGCGTCTGATTTGGCTCTAAATCGTCACAGCATGGACGACGAGATTAATCAACCTGA	244			
188	GTGAGATTAATTTGGTCTTAAATCGTCATTCATGATGATGATGACGAGATTAATTAAGCTTA	247			
245	ACCAGCTCTCGAAAAGTGAAAGCCGCGCCAGGTTACTGTTGGCGTAATCCGACACCG	304			
248	ACCAGAGTTTGGAAAAGTGTAAGGACGAGCGCAAGTGACAGTGGCGTAAATCCAGATACG	307			
305	GTAAACGACTACTCTGTGGAAACCGGTTCCGACCTACGACGGTTTACAAACAACGAAACGC	364			
308	GCAATATGTTATATGTTGTGGCCAGTGCGTACTTACAGGCTTACATTAATGAATGAAGACGC	367			
365	GTACAGCAGTACTCCGCGAATTTTCAGCAGAAAGCATGATTCGACAGTGCAGAAACAGGAA	424			
368	GCCACAATATTTGTGAGAAATTTTCAGCAAAAGCGCATGATTTCCAGGCGACGAAGCAGAGA	427			
425	TCTAGCGCACCGCGTGCCTCAGCCGGATGCGCGCTGGCAGGTGATTTAGCAGCGAAGAA	483			
428	TTTACGGCAGCTCATCCGCGCAACCGGATGTGCTGGCAAGTCAATTCACAGCAAGAA	486			
RESULT 2	AF184152	4983 bp	DNA	linear	BCT 01-OCT-2000
LOCUS	AF184152				
DEFINITION	AF184152				
ACCESSION	AF184152				
VERSION	AF184152.1				
KEYWORDS	GI:10441343				
SOURCE	Piscirickettsia salmonis.				
ORGANISM	Piscirickettsia salmonis.				
REFERENCE	1 (bases 1 to 4983)				
AUTHORS	Kuzyk, M.A., Burian, J., Thornton, J.C. and Kay, W.W.				
TITLE	Identification of a genus-common Rickettsial surface antigen in the				
JOURNAL	salmonid pathogen Piscirickettsia salmonis				
REFERENCE	2 (bases 1 to 4983)				
AUTHORS	Kuzyk, M.A. and Kay, W.W.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-Sep-1999) Biochemistry & Microbiology, University of				
FEATURES	Victoria, Pech Building, PO Box 3055, Victoria, BC V8W 3P6, Canada				
SOURCE	Location/Qualifiers				
	1. 4983				
	/organism="Piscirickettsia salmonis"				
	/strain="LF-89"				
	/db_xref="taxon:1238"				
	/clone="PB12"				
	<1. 501				
	/gene="a1r"				
	<1. 501				
	/gene="a1r"				
	/codon_start=1				
	/transl_table=11				

gene  
CDS  
complement(526. .1257)  
/gene="bax"  
/gene="bax"  
/codon\_start=1  
/transl\_table=11  
/evidence="not experimental"  
/product="BAX"  
/protein\_id="AAGI7002.1"  
/db\_xref="GI:10441348"  
/translation="MSLAIVITAYLAQOPLFSAPTSTQINTAVSODIOSIKOEFTIHK  
KTKVSTNITQIILQKROHILSLQKRSLEQITQLQAKQKLNHLHKHPQDS  
LLIKRDIIPNALISAQAINESMGSRPEAVEGNNEPCMRHYGGCITIPARANNH  
WNAVNSWTASVRAVYHILNTHNAVQALRDLRAIMRANHODVSAFKLBSLTAVSIK  
GKRYELIOTITTRLKESISQIP"  
1373. .2743  
/gene="rada"  
/evidence="experimental"  
1373. .2743  
/gene="rada"  
/codon\_start=1  
/transl\_table=11  
/product="DNA repair enzyme Rada"  
/protein\_id="AAGI6998.1"  
/db\_xref="GI:10441344"  
/translation="MSKQNOYVCSDCGGIAKKMLWGPHCOQKNSLSEVKEYILPNRP  
GNSQAEAFAGIDAPKVCLESEITPEQISRPQVCEIDPRYLGGIVHGAVILIGRP  
GIGKSTLITLQSVNCTQKGLVLYVTGESLEQVTLRSKRIGSLSDVDLRLATQYER  
ILKAEILQKPYLIYDSIQITPTFSLOSAPEGVAQVRESAIILIQFAKRTGCLFLVG  
HYTKGALAGPVEHEHWDTVLYFEGEDSREFLRAVKNFEGANBELGIFAMTEGL  
KTVNSPSAIFLSREDDLPQSVVWVAMEGTRPLVEYQALVDESHSNPRTAVGQD  
QRLAMLVLAVNRHGIAISYDDQVFINVVGCMKIFETAAIDLALLACVSSLRGALSOE  
LIVFEIETLSEIRPVGRLERIKEAQVHGFTQALVLANCPKQKVDGIEIVGKHLE  
DALATL"  
2834. .3322  
/gene="ospa"  
/evidence="experimental"  
2834. .3322  
/gene="ospa"  
/note="Ssimilar to the rickettsial 17 kDa antigen"  
/codon\_start=1  
/transl\_table=11  
/product="17 kDa antigen"  
/protein\_id="AAGI7000.1"  
/db\_xref="GI:10441346"  
/translation="MNRGCLQGSLLIITSVELVCAQNFPSROEYGAATGAVVGGVAAQ  
LFGSGSGVMAAIGAVLIGLIGSKTQSDMDQDKIKLNSLEKRAQGVTRRNPDT  
GNSYSEVEYRIQRYKNQERRQCYCREFOQKAMLAGCKQKELTYGACRQPDGRNVIYST  
EK"  
3560. .3846  
/organism="Piscirickettsia salmonis"  
/strain="LF-89"  
/db\_xref="taxon:1238"  
/clone="pb12"  
/insertion\_seq="insertion sequence-like element"  
3560. .3846  
complement(3842. .4561)  
/gene="tnpa"  
/complement(3842. .4561)  
/gene="tnpa"  
/codon\_start=1  
/transl\_table=11  
/product="transposase"  
/protein\_id="AAGI6999.1"  
/db\_xref="GI:10441345"



```

/translation="MDTLISLFCV>IDDFCQELL>PQWMAILLIEDNKRKRNPQOMSTN
EIIMTIVIFPKSNVREKMYLHVINGSWAKFPNVSYNREVELMPSILPLLOCFIA
AQGTANGIYFVSDTILRVCHERHARQONAFGLAKSKSTNGVYIGKKLIIYNDMG
ELMAFKSKSTQDTRDVLVPMANLNLGKIIIGQVYSIKTLEPQJLTXEKLITITIRKN
MKNLVIMIDKILLKRGIIA"

```

repeat_region	4562..4849	
BASE COUNT	1362 a	778 c 1265 g 1578 t
ORIGIN		

Query Match	61.1%;	Score 295;	DB 1;	Length 4983;
Best Local Similarity	76.0%;	Pred. No. 2e-61;		
Matches 364;	Conservative	0;	Mismatches 115;	Indels 0;
			Gaps	0;

QY	5	GTGGTGGCTCCAGGGCAGCTCTGCATCATTAATCTCGTTTCCGTGGTGGTCCGCC	64
Db	2841	GAGATGTTTCCAAAGTAGTAGTCTAAATTAATTATACGTGTTTTAGTTGGCTGTGCC	2900
QY	65	AGAACTTCAGCCCGCCAGAGAGTTGGCGCGCCACCGTGCGGTTGTGGCGTGTGGCG	124
Db	2901	AGAACTTTAGTCGTCAAAAGCAGCGAGCTGGGACGTGGGGCTGTGTTGGCGGTGTGGTG	2960
QY	125	GCCAGCTTTTGGGTAAAGCGCTGTGGTCTGTGTGTCATGTGGCCATGGCGGGTGGGTTGCG	184
Db	2961	GCCAGCTTTTGGTAAAGTAGTAGTCCAGTTGCATGTGCCATTTGTGTGCTGTTTTGG	3020
QY	185	GCGGTCGTGATTTGGCTCTAAATTCGGTTCAGAGCATGAGCAGCAGAGATAAATCAATCGA	244
Db	3021	GTGGATTAATTGGTCTTAAATCGGTCAATGATGATGATCAGAGAGATAAATTAAGCTAA	3080
QY	245	ACCACTCTCTGGAAAAGTAGAAACGCCGCCAGGTACTCTGTTGGCGGTAAATCCGGACCCG	304
Db	3081	ACCAAGCTTTGGAAAAGGTAAAAACGACGGGCAGAGTACACGTTGGCGGTAAATCCAGATACAG	3140
QY	305	GTAACAGCTACGCTGTGGAAACCGTTGTCACCTACACAGGTTACAAACAACAGAAACCC	364
Db	3141	GCAATAGTTAATAGTCTTAAAGCCAGTAGTGGCTACTTACCAAGCTTACAAATTAAGCAAGAGCGTC	3200
QY	365	GTCAGCACTACTGCCGCGAATTTTCAGCAAAAAGCCATGATCGCAGGTGAGAAAACAGAAA	424
Db	3201	GCCAGCATTAATTGTGCAAAATTTTCAGCAAAAAGGCGATGTTTTCAGGCGCAAGAAAGCAAGAGA	3260
QY	425	TCTTACGGACCGCGTCCCTCAGCGCGGATGGCCGCTGGCAGGTGATTAAGCACCGCAAAA	483
Db	3261	TTTTCAGGCACTGCAACGCGGACCGGATGTGCTGTGGCAAGTCAATTTTCAACAGCAAAAA	3319

RESULT 3	LOCUS	537 bp	DNA	linear	BCT 04-FEB-1999
R1R17KGA	R1R17KGA				
DEFINITION	Rickettsia japonica	17k genus-common antigen gene,	complete cds.		
ACCESSION	D16515				
VERSION	D16515.1	GI:695412			
KEYWORDS	17k genus-common antigen; 17k dalton protein.				
SOURCE	Rickettsia japonica (sub-species: YH)	DNA.			
ORGANISM	Rickettsia japonica				
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.				
AUTHORS	1 (bases 1 to 537)				
TITLE	yoshida, Y.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (03-JUL-1993)	yoshiya yoshida, Kanagawa Prefectural			
AUTHORS	Pabulic Health Laboratory; 52-2 Nakao-cho, Asahi-ku, Yokohama,				
TITLE	Kanagawa 241, Japan (Tel:045-363-1030, Fax:045-363-1037)				
REFERENCE	2 (bases 1 to 537)				
AUTHORS	Furuya, Y., Katayama, T., Yoshida, Y. and Kaiho, I.				
TITLE	Specific amplification of Rickettsia japonica DNA from clinical				
JOURNAL	specimens by PCR				
MEDLINE	J. Clin. Microbiol. 33 (2), 487-489 (1995)				
FEATURES	95229950				
source	Location/Qualifiers				
	1..537				
	/organism="Rickettsia japonica"				
	/sub-species="YH"				

```

/db_xref="taxon:35790"
/note="propagated in Vero cells"
58. .537
CDS

```

BASE COUNT ORIGIN	176 a	92 c	129 g	140 t
----------------------	-------	------	-------	-------

Query Match	9.68;	Score 46.6;	DB 1;	Length 537;
Best Local Similarity	47.28;	Pred. No. 0.5;		
Matches 176;	Conservative 0;	Mismatches 194;	Indels 3;	Gaps 1.

Oy	100	GATGGGTTGTTGGGGGTTTCCGGCCAGCTTTCGGTAAAGGCTCTGTGTGTCG	159
Db	163	GCGGTGCTGGTGGCGCATTACTTGGTTCATTTCCGTAAGGGCCACAGACACTTGTG	222
Oy	160	ATGGCCATTCGGCGGTGCGGTTCTGGGCGGTCTGATTGGCTCTAAATCGGTCAGACATG	219
Db	223	GGAGTAGGTGTAGTGTCAATTACTTGGAGACGTTCTTGGTGGACAAATCGGTCCAGGTATG	282
Oy	220	GACCAGCAGGATAAATCAAACTGAACCACTCTCTGGAAAAAGTGAAGCCGGCCAGTT	279
Db	283	GATGGACAGGATAGAAAGACTTGCACAGGCTTACCTCACAGAGAGCTTTA--GAAACAGCT	339
Oy	280	ACTGTTGGCGTAATCCGACACGCGTAACACTTACTCTGTGGAAOCGGTTGGCACTTC	339
Db	340	CCTATGTGTAGTACGTAGAAATGCGCTAATCCGGATAACGGCAATTAACGGTATCGTAAACA	399
Oy	340	CAGCGTTACAAACAAGGAAACGCGCTCAGCACTACTGCGCGGAATTTACAGCAGAAAGCC	399
Db	400	CCTAATTAACCTTAATAGAAATAGCACTGGTCAATATTGGCGTAGACTACACTCAAAACAGTT	459
Oy	400	ATGATCGCAGGTCAGAAACAGGAAATCTAAGCGCACCGCTGCCCTTCAGCCGATGGCGCC	459
Db	460	GTAATATAGCGGAAAAACAACAAAAACATACGGTAATAGCTGCCGCCAACCTGACGACAA	519
Oy	460	TGGCAGGTGATTA	472
Db	520	TGGCAAGTTGTGA	532

RESULT 4	
RPX04	
LOCUS	
DEFINITION	RPX04 237523 bp DNA linear BCT 11-NOV-1998 Rickettsia prowazekii strain Madrid E, complete genome; segment 4/4.
ACCESSION	AJ235273 AJ235269
VERSION	AJ235273.1 GI:3861237
KEYWORDS	complete genome.
SOURCE	Rickettsia prowazekii. Rickettsia prowazekii. Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.
ORGANISM	1 (bases 1 to 237523) Anderson,S.G., Zomorodipour,A., Andersson,J.O., Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K., Eriksson,A.S., Winkler,H.H. and Kurland,C.G. The genome sequence of Rickettsia prowazekii and the origin of mitochondria
TITLE	Nature 396 (6707), 133-140 (1998)
JOURNAL	99039499
MEDLINE	2 (bases 1 to 237523)
REFERENCE	Anderson,S.G.E.
AUTHORS	Direct Submission
TITLE	Submitted (11-NOV-1998) S.G.E. Anderson, Stv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN

FEATURES  
source  
Location/Qualifiers  
1. .237523  
/organism="Rickettsia prowazekii"  
/strain="Madrid E"  
/db\_xref="taxon:782"  
164. .1123  
/gene="Rp701"  
164. .1123  
/gene="Rp701"  
/codon\_start=-1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA15137.1"  
/db\_xref="GI:3861238"  
/translation="MAEPIITLLFGKRIIDLRIRNOKALNKLTHTPILOVNESHS  
DKERISEILEKSYKECKIPITFINQISNNNAKPLFRADOLINQONLKITITFDKN  
ALPONSLEKSYWQOIERVDHVEFTNEADNLSIADIIPKDKTTITDRLVTSFN  
NLVYHETNGLSTIYKVKDKLITKAKNODSKEIETPMSVDEATNLITAFGLTS  
EDQIYGKLEIHEILKDPNNAENFKKYSQISQPKDKDKDKDKNIDSNFTQKVI  
NDSNKDIOVEHITSYVODKEKQSPQGLRIRINRYFNKINYEKQ"  
2074. .3348  
/gene="Rp702"  
2074. .3348  
/gene="Rp702"  
/codon\_start=-1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA15138.1"  
/db\_xref="GI:3861239"  
/translation="MTTIVIVIIIMISLSALFAETATATASSPKRIKLKTAGNKR  
AKTVLEVLKKKEKRVIGTLIGNSLINICTTIAITLLEISLGNNGTIVASGVAFTI  
VFAEVVKAJAVAKPEQLTKMSTVIYFLKRPINIALDYITKICFPFRINLQ  
ISGEVEYRGVIEHHOEGVYKSERNMLGILDIRMTVSEIMTHRSNITAILNDLPY  
ELIKTLTSAHRIRIPLOMDNRNIGIILKLLKALYENNNDDKVDINTLTPPM  
FITDNLAVDQLHAFERNNHPCVDEIGTLGITLEDVIEIVGPTIDEHRLN  
EIIQESNTEFIIKGTITIRDIRNELDMNLSDEDANTAGLIHKIARIPOGEVIKIF  
NFKIILKKIANKIDSVKITVLPMTETTISSE"  
3341. .5353  
/gene="Rp703"  
3341. .5353  
/gene="Rp703"  
/codon\_start=-1  
/transl\_table=11  
/product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF  
(comp)"  
/protein\_id="CA15139.1"  
/db\_xref="GI:3861240"  
/translation="MSKIGNFLLLTMCISGILTLIIIPWLDYLKKRHRNNICLKEF  
PMIPCTSOVMTQAITFEYASSISIFAFCULVAFITISDFVQNVFLHSSTLKPLIKI  
SGSMSSHESGMLLMFCILQIVSCYIECEDMKMKFSIIILSAIOLFFSFIYFSN  
PHEMFSPVKEGILNPMLODIALSIHPLLYIGVYVAAPTIVCAIFLIAVISRLD  
NMLPGVTCYKTIQSLSDRTISIRIKIFACGILFTTIGISLSMWARBLNNGLMF  
PDVENISLSPMLSGTMLHSITVTRKRMQHTIILSITFLVIFSTFLVRSGET  
NSIHAFASSESRMIVLYILCIIGIGLVYIIRLNCSIEKTIYNTFNMTSHSCNMT  
IKKAILIGNIFELSLILICATLYPTIYALFNDKPIIINEEFIKNFIVIPILIC  
TILEFTTNSIKIKHIIIMLSLITIVYILNIKFSVISIITIMTSIFMLHNYHLYL  
KTYVFRNSLKASTSMILGHEGALLAFSTNMSLQSEMDPTGEGVSTFNPAFKYT  
LQNIKFAQGNKYROIJAEEFLBDHKNNITILKRPENRYITTEOKLSQSDSHSLYLDL  
YALVSNIDKIIIAKTIYKPMASFIMIGVILTTAGVIALIRKNSY"  
7264. .12195  
/gene="Rp704"  
7264. .12195  
/gene="Rp704"  
/codon\_start=-1  
/transl\_table=11  
/product="CELL SURFACE ANTIGEN (scat)"  
/protein\_id="CA15140.1"  
/db\_xref="GI:3861241"  
/translation="MAOKPMFLKKIISAGLVTPASTATYVAGSGVAMGAMQYRNTTN  
AAATPDGIGEDQAGANIPVAPNSVITANANPNFENPENGSLNSIFLDTADLAVT  
INDDTLIGFTITTAQOKAFNFYVAAGKILINITGGGITYVEASNTITNAQALIKVHG  
AAINANDLSGISITFAAPASVLEFNLINITGDEAPLTGLGANSKIVNGGGLNTING  
FIVQSNTEFAGIKTINIDDOGLMFNSTPDANLTNLQVGNNTINFGIDGTGLVLY

SKNGAATEFNVTGILGNLKIIELNTPAAVAGKILISOGAANAIVIGTNGCAGRAGPI  
VSYNDGMAATISGOVYAKNMVITOSANAGGVFEEHIVDVGITTEKPKDSVITTE  
NSMFGSTNFGENDTOIVYPTKILKEFIDQYKRNNGTAVITFMNAGALYSSTDPN  
IATVNTALIEBAGVVEISGIIHAEIRKJFIDQYKRNNGTAVITFMNAGALYSSTDPN  
LAAGSIQDGSAAITTDIGNGVNAALOHITVANDSKIALDGMANIGANVGAIAHE  
QANGGIKLTNTQNNIYVNFEDDITTDGTGVNDASSLTNNQTLINGSISTVANTKT  
LAQNTIGSSKTIINAGDVAINEIVENGSVOLNHNRYLTITKTITANAGQOLIYAADP  
LANTNTLADGTNINSAENPLSTIHPATKANADSIINVGKVALYNNITNTNANGS  
LHFRSGSTISVSTVGQOCHKLNNLIDNGTYKPLGDTTFNGGKRIEBSKLIQJLSN  
NVTTHVESADNTGTLEFVNTDPIYVTLNKGAVFGYKQVITISGFNIVFNEITGVG  
IVHGILANSISFENASLGTSLEPSPGLDVLTKSVNGVATDNNAPLVVVGAGIDS  
MINNGOIIIGDKKNIILALSIGSDNSITVYNAATLSGRTTNNNGTVALTSGMPPNNGT  
ITGIGLENGSPKIKOVYFTTDYNNLGSIIINANTYINDYVLTGAGTGFQDAKITLG  
SVNGANVRPVDSPDPRSMIVATQANKCTVYIGAPLVSNTGSLDTPYASVRFNGN  
DSGAGIQRNTYSNIDFGYNTLINSNTVILSGGTTAINEIDLLTNLINFANGSTW  
GNDTISTITLNTVSSNGIGOVYIAEDQVNAATITGTTTIRKIQDANANFSGTQAVTLIQ  
GGARFNGTILGAPFAVYGSNIPEYELIRSDNDVYLTFRNDLVANVTYVNAVGAISAN  
APGVSONISCELESTMTAAVNNMLAKPDSVATFPGALITDPSAAVTTVNAEDQTKG  
ODLSNRLGTLRYLSNAETSDVAGSAGVSGDEAVSVGWAKFPYVNAEDQKKG  
IAGYKAKTGTGVVYGLDTIADNDIMIGAIGITTDIKHODYKKGDKTDINGLSFLYG  
SQOLYKRNFAQGNATITLTKVKRSQRPYFESGKRSKOIAAGNYNMPGSLITGXY  
DINAMPVILVTPMAGLSYLSKSNENKTETITVANKRINRSDDRDVLGAVAGVST  
VNTDVIYIPEIHSFVHKVNGKLSNSQSLMDQTPAFISQPDRTAKTSYINGLSANI  
KSDAKMEYIGIGYDFNSASKYTAHQGTLKVRVNE"  
12953. .13216  
/gene="Rp705"  
/note="Pseudogene that lacks ATG or GTG initiation codon.  
Homology to guanosine-3,5-bis(diphosphate)  
3-pyrophosphohydrolase."  
12953. .13216  
/gene="Rp705"  
/complement(12953. .14507)  
/gene="Rp706"  
/complement(13563. .14507)  
/gene="Rp706"  
/complement(13563. .14507)  
/gene="Rp706"  
/codon\_start=-1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA15141.1"  
/db\_xref="GI:3861242"  
/translation="MKIRQYAKPVYIIGSPBELTDAERLEFEEHNPGLITLPRNRIRK  
NEKGEQDKAALIKLADIKEVLGNDNTIADIDGQVAKRLIAPFYDAPAPQFTBLQ  
TCKYNTSTIKELREVGINDPFAVDLHDGDAKILISDSKFESEIYVFLPSAIA  
GLQEEVYTAICRKHIPGHRATVSHLEPIIDGSLTLEPTDFKVRKELAKYDIKIA  
MTAHVYITADPDNPATLSQIYDIYKSNIGFGLIISDPTIEMQALSGSMADITKRAL  
DAGVDIVLEGTGEPNKMSEYLSVAERSIDKFSDFLTS"  
14552. .15025  
/gene="Rp707"  
/complement(14552. .15025)  
/gene="Rp707"  
/codon\_start=-1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA15142.1"  
/db\_xref="GI:3861243"  
/translation="MINNKLTKRYVASEVYRILNLTALHKRLRYLETIKDGLSHYKINN  
RKYVTAHDIDLOKPLFVSREREYLVRCYGRHSRDXDYLANFSNHGASDPSRQNEOG  
KNSCEITINHVOORQOSQAVYVSLMDNVTSSRIDLTTFNHLISIQIKKILLIS"  
15220. .15507  
/gene="Rp708"  
/complement(15220. .15507)  
/gene="Rp708"  
/codon\_start=-1  
/transl\_table=11  
/product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (hima)"  
/protein\_id="CA15143.1"  
/db\_xref="GI:3861244"  
/translation="MTITKNIKISLMSLSKLGFSNNCEAIIVNTVFSNILEIAVQKLT  
LKNFGSEYKQKPPRGINHTKSPVWASKKMLRTPSPSKKALINKSML"  
18405. .18701  
/gene="Rp709"  
18405. .18701  
/gene="Rp709"

[illegible]

ORANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	
Ralstonia solanacearum	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;	1 (bases 1 to 203050)	Submitted	Direct Submission	
		Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Ariat, M., Billault, A., Brottier, P., Camus, J. C., Catolico, L., Chandier, M., Choise, N., Cladel, Renard, C., Cunne, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Siguer, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. and Boucher, C. A.	Genome sequence of the plant pathogen Ralstonia solanacearum	Unpublished	
		2 (bases 1 to 203050)	Submitted	Direct Submission	
		Boucher, C. A.	Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Molculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMG CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex	http://seqence.toulouse.inra.fr/R.solanacearum.html	
FEATURES	Location/Qualifiers				
SOURCE	1. 203050				
	/organism="Ralstonia solanacearum"				
	/strain="GM11000"				
	/db_xref="taxon:305"				
gene	complement(180..1025)				
	/gene="RSC2606"				
	/note="RSC0854"				
CDS	complement(180..1025)				
	/gene="RSC2606"				
	/function="elements of external origin; plasmid-related functions non-bacterial functions"				
	/note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed"				
	/codon_start=1				
	/transl_table=1				
	/evidence="not_experimental				
	/product="HYPOTHEICAL REPLICATION INITIATOR AND TRANSCRIPTION REPRESSOR PROTEIN"				
	/protein_id="CAD16313.1"				
	/db_xref="GI:17429628"				
	/translation="MSSPGGALPQRQQLDLFRALPCDMMPRSDLMARPPPSLAKS RRTADIDFRAGGITRVEGTREHETAIITWADVLIWASOIVEARDNGIIPRSWIRAT PYELRPIGRSTSLNDYQRLKALDRQSTASITSJETGRRLHRPSWINENELAD ASGPFELIILIPMEYAGVDALVLTIDPAYFRILGIEIRMLYRLVRHGRGROENG WQPFKFLHRRKSSAARNSPFDYDLRALVARQSLPGYVLGIERMPDGGELIFRPVL HTRAG"				
gene	complement(1052..1327)				
	/gene="RSC2607"				
	/note="RS00855"				
CDS	complement(1052..1327)				
	/gene="RSC2607"				
	/function="miscellaneous; hypothetical/partial homology"				
	/note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed"				
	/codon_start=1				
	/transl_table=1				
	/evidence="not_experimental				
	/product="HYPOTHEICAL PROTEIN"				
	/protein_id="CAD16314.1"				
	/db_xref="GI:17429629"				

gene . . . . . 2212

/translation="MRPAPLRPAAPAPQPORRYLTLPNPAEYLRSPRTLEKQKVVGS  
GKRFKRGRRVYAVSDLDNADQNSYATSDPEYAEHNAHDYDGR"  
complement(1442. . 2212)  
/gene="RSC2608"  
/note="RSC2608"  
complement(1442. . 2212)  
/gene="RSC2608"  
/function="miscellaneous; hypothetical/partial homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="HYPOTHETICAL/UNKNOWN PROTEIN"  
/protein\_id="CAD16315.1"  
/db\_xref="GI:17429630"  
/translation="MTDRSAEHWYPTAAVLYLVHLIDGPAALWEI LRKNPDYRRMILRR  
RRRPPDAQWGLRLIEDPALDRADHPAMPFPDHDVVOLFPDADPEEHAHFEWRVP  
GKKOLIHDKRLVIVSHWPGCLRLALAGLEDMAVLYATRCATPCARVRLTAL  
DLALATVYAAPTPAARSPTLAAVLELHTLQALDATALGASLRVAVGLFSADAVAAD  
WKKDSLRRVRYLRVRCGBALMRGGRRLQLPILPQ"  
2213. 8108  
/note="ACUR or RS00937  
predicted by Framed"  
/evidence=not\_experimental  
complement(2557. . 3150)  
/gene="RSC2609"  
/note="RSC2609"  
complement(2557. . 3150)  
/gene="RSC2609"  
/function="miscellaneous; hypothetical/partial homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon\_usage"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="PROBABLE LIPOPROTEIN"  
/protein\_id="CAD16316.1"  
/db\_xref="GI:17429631"  
/translation="WTERPLSCFTYGNPLRLPTPEGNRIRVFPASDNSASVLPFGCC  
DLHILRASTRMRHRSYIVLCFTGRAPRIIVGFVRLLPPTLPVLADEARLARLHAGAL  
AAGORIDLPVRLFTPDAAHVWLVLSLDPDGDTAYGLIDIGTGLPSLTGIVKLSDLA  
GIVGPKRLPEVRDRYFOAAHLLSEYVRLAENGAITD"  
3223. 3513  
/gene="RSC2610"  
/note="RSC2610"  
/note="RS00859"  
3223. 3513  
/gene="RSC2610"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD16317.1"  
/db\_xref="GI:17429632"  
/translation="MAAKYSLAKALKTVRKARGLSOEAFSDVSSRTYMSLTIERPLKSP  
TLHKLTICEVCEVDVHPLTLTLTAFAGDSTRADELLAQVRELAVINSDSD"  
3522. 4085  
/gene="RSC2611"  
/note="RSC2611"  
/note="RS00860"  
3522. 4085  
/gene="RSC2611"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage"

gene . . . . . 2212

/translation="MRPAPLRPAAPAPQPORRYLTLPNPAEYLRSPRTLEKQKVVGS  
GKRFKRGRRVYAVSDLDNADQNSYATSDPEYAEHNAHDYDGR"  
complement(1442. . 2212)  
/gene="RSC2608"  
/note="RSC2608"  
complement(1442. . 2212)  
/gene="RSC2608"  
/function="miscellaneous; hypothetical/partial homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="HYPOTHETICAL/UNKNOWN PROTEIN"  
/protein\_id="CAD16315.1"  
/db\_xref="GI:17429630"  
/translation="MTDRSAEHWYPTAAVLYLVHLIDGPAALWEI LRKNPDYRRMILRR  
RRRPPDAQWGLRLIEDPALDRADHPAMPFPDHDVVOLFPDADPEEHAHFEWRVP  
GKKOLIHDKRLVIVSHWPGCLRLALAGLEDMAVLYATRCATPCARVRLTAL  
DLALATVYAAPTPAARSPTLAAVLELHTLQALDATALGASLRVAVGLFSADAVAAD  
WKKDSLRRVRYLRVRCGBALMRGGRRLQLPILPQ"  
2213. 8108  
/note="ACUR or RS00937  
predicted by Framed"  
/evidence=not\_experimental  
complement(2557. . 3150)  
/gene="RSC2609"  
/note="RSC2609"  
complement(2557. . 3150)  
/gene="RSC2609"  
/function="miscellaneous; hypothetical/partial homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon\_usage"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="PROBABLE LIPOPROTEIN"  
/protein\_id="CAD16316.1"  
/db\_xref="GI:17429631"  
/translation="WTERPLSCFTYGNPLRLPTPEGNRIRVFPASDNSASVLPFGCC  
DLHILRASTRMRHRSYIVLCFTGRAPRIIVGFVRLLPPTLPVLADEARLARLHAGAL  
AAGORIDLPVRLFTPDAAHVWLVLSLDPDGDTAYGLIDIGTGLPSLTGIVKLSDLA  
GIVGPKRLPEVRDRYFOAAHLLSEYVRLAENGAITD"  
3223. 3513  
/gene="RSC2610"  
/note="RSC2610"  
/note="RS00859"  
3223. 3513  
/gene="RSC2610"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD16317.1"  
/db\_xref="GI:17429632"  
/translation="MAAKYSLAKALKTVRKARGLSOEAFSDVSSRTYMSLTIERPLKSP  
TLHKLTICEVCEVDVHPLTLTLTAFAGDSTRADELLAQVRELAVINSDSD"  
3522. 4085  
/gene="RSC2611"  
/note="RSC2611"  
/note="RS00860"  
3522. 4085  
/gene="RSC2611"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage"

```

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAD16318.1"
/db_xref="GI:17429633"
/translation="MAYRNGNYTAFVYAPPEPSSLGANATKDFQYNTLRMMKGADA
TFPVDSHDRTYSVRDGSMDREYTLKPLRLERLRSSKNIIILISITTSNARLREIDY
GINDQGLPIVITYPEVATRKESLDVDGSLQAVKMLMGNLPEVFQDSMSKIPTLVHPMK
ELIKRSJDEFFWMSGKAPDIYRYNP"
4107..4955
/gene="RSC2612"
/note="RS00912"
4107..4955
/gene="RSC2612"
/note="RS00912"
4107..4955
/gene="RSC2612"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD16319.1"
/db_xref="GI:17429634"
/translation="MIHVMAKVNFEDKRVIKKFLEITAVVSGLSIFIVIEVDPAEKK
LKADIAFLAIALIYYVVYLWSNNLSINDINESSDYAIRTGDIFOOPGAKALAFNY
FDTVDNKRIIGETLVNGVFIQKHGVPSIELDRYEYAESEVELENBARKYGKKK
RVQIQTICLYKDYLITAFSPKDENNKALLIMPEYLEPLINDKVNVAVOQSSTIIR
SGSITTRIKGHKNISDEDLKLIMLTFRISMRRKRYAKLIYIHKKRIDQINLDLIK
SARGNV"
complement(5305..5619)
/gene="RSC2613"
/note="RS00913"
complement(5305..5619)
/gene="RSC2613"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="CONSERVED HYPOTHETICAL PROTEIN"

Query Match      8.9% Score 42.8; DB 1; Length 20350;
Best Local Similarity 56.3%; Pred. No. 4.7;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY   327 GGTTGGCAACTACAGCGCTAACAAACAGGAACGACCGGTACGACGTACTGCCGCAGATT 386
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105900 GCTTGCCGTCTGCGCCTGCGCATGCGCAGGACATCCCGGTGCGCACTGCGGCCGAT 105959

QY   387 TCAGCAGAAGGACATGATCGAGGTACAGAAACAGGAATCTACAGCACCGGTCGCCCTCA 446
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 105960 GGTGGGGGACAGCGCAGATCGACGGGGTAGTGACAGGTACAGCGCACGCGGTGCTGCA 106019

QY   447 GCCGATGGCGCGCTGGCAGGTG 468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106020 GCCGACGGCACCTTGCGCAGATTG 106041


RESULT 6
LOCUS       RSU76907               448 bp    DNA             linear     BCT 18-FEB-1998
DEFINITION  Rickettsia sp. 17kDa common-antigen gene, partial cds.
ACCESSION   U76907.1 GI:2894794
KEYWORDS
SOURCE      Rickettsia sp.
```

	ORGANISM	Rickettsia sp.
	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;	
	Rickettsiaceae; Rickettsiae; Rickettsia.	
	REFERENCE	1 (bases 1 to 448)
	AUTHORS	Davis,M.J., Ying,Z., Brunner,B.R., Pantofila,A. and Ferwerda,F.H.
	JOURNAL	Rickettsial relative associated with papaya bunchy top disease
	Curr. Microbiol. 36 (2), 80-84 (1998)	
	MEDLINE	98087556
	REFERENCE	2 (bases 1 to 448)
	AUTHORS	Ying,Z. and Davis,M.J.
	TITLE	Direct Submission
	JOURNAL	Submitted (01-NOV-1996) Tropical Research and Education Center, University of Florida, 18905 SW 280 Street, Homestead, FL 33031, USA
FEATURES	source	Location/Qualifiers 1..448 /organism="Rickettsia sp." /db_xref="taxon:789" CDS <1..>448 /note="Puerto Rico" /note="putative" /codon_start=3 /transl_table=11 /product="17 kDa common antigen" /protein_id="AAC02809.1" /db_xref="GI:2894795"/translation="ALASMLEACQPGFNKKDGGTLIGTLLGLVSGFGGCGTRIAAVAGALLAILGNQIAGMDIEDGRKLAELTSORALEAAPSGSVVVRNDRNDNGYTATPSKAYKNNTGOVCREFYQTGYVYGQQAKYGTACNPDSQMNV"
BASE COUNT	135 a 81 c 120 g 112 t	
ORIGIN		
Query Match	Best Local Similarity 47.7%; Score 41.6; DB I; Length 448; Matches 184; Conservative 0; Mismatches 199; Indels 3; Gaps 2;	
OY	87 TGGCCGCCACCACCGCTGGTGTTGGCGGCGTTCCGCCACGCTTCGGTAAGGCCTC 146 	
Db	65 TGCTACGCTATTGGCGGAACCTTAGGAGAAGACTTGTGTTCAACAATTTGGTGGGTAC 124	
OY	147 TGGTGTGTCGATGGCATCGCGGGCGGGTTCGGCGGCTGATGGCTAAAT 206 	
Db	125 GGAGACACTTGTCTGGGTAGACAGCACAGTGCTTCTGTGATCTTTGGCAACCAAT 184	
OY	207 CGGTACAGACATGCACACAGATTAATCAAATACTGAACACAGTCTTGAAAAAGTGA 266 	
Db	185 TGGTCAGGTATGATGTAACAAGATGAAGAAACTTGCACAGCTTACTTCCAAGAGCTT 244	
OY	267 AGCCGGCCAGGTACTGCTTGGCGTAATTCGGACACCCGGTAACAGCTACTGTGGAAC 326 	
Db	245 AGA-AGCTGCACCAACGCGTAGCACTGTACATGCGCTAATCTCATATGTAATTATG 303	
OY	327 GGTTCGCACCTACACGCGTTACAACAAAAGSAAGCGCGTACGACGTACTGCCGGGANT 386 	
Db	304 GTACTGTACACCAAGCAAGAAAGTTATAAATAATAATACCG-GTCAATATTGTCGGAATA 361	
OY	387 TCACGAGAAGCCATGATCGCAGTCAAGAACAGAAATCTACGCGACCGGTGCCCTCA 446 	
Db	362 CACCCAAACAGTGTAGTAGTGGAACAAACAAACAAACTTATGCACTGCTGCCGTCA 421	
OY	447 GCCGGATGGCGCGTGCAGGTGATTA 472 	
Db	422 ACCTGATGACAAATGCAAGTTGTGA 447	
RESULT 7		
LOCUS	AL646061	198050 bp DNA linear BCT 07-DEC-2001
DEFINITION	Ralstonia solanacearum GM1000 chromosome, complete sequence; segment 5/19.	
ACCESSION	AL646061 AL646052	
VERSION	AL646061.1 GI:17427781	
KEYWORDS	.	

SOURCE	Ralstonia solanacearum.
ORGANISM	Ralstonia solanacearum Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
REFERENCE	1 (bases 1 to 198050)
AUTHORS	Salanoubat,M., Genin,S., Artiguenave,F., Guzy,J., Mangenot,S., Ariat,M., Billault,A., Brotlier,P., Camus,J.C., Catalicio,L., Chandler,M., Choinsé,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Lavié,M., Moisan,A., Robert,C., Saurin,M., Schlex,T., Siquier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A.
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 198050)
AUTHORS	Boucher,C.A.
TITLE	Direct Submission
JOURNAL	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMCM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA UMGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex Christian.Boucher@toulouse.inra.fr http://sequence.toulouse.inra.fr/R.solanacearum.html.
COMMENT	Location/Qualifiers
FEATURES	1..198050
SOURCE	/organism="Ralstonia solanacearum" /strain="GM11000" /db_xref="taxon:305" 151..642 /gene="RSc0770" /note="RS05084" 151..642 /gene="RSc0770" /EC_number="1.-.-.-" /function="small molecule metabolism" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (SMALL SUBUNIT) OXIDOREDUCTASE PROTEIN" /protein_id="CADI4472.1" /db_xref="gi:17427782" /translation="MHTRPPSSDDDDPAGLTRROWLSALATATASVLTLYAVNAQ APEPELDAPFMGLSGLTRRSALDRAYACARLLAAQKASAGFAAORPLAQSILAGTISL GLQGERLRLTIQAMTIGVGVDTVTYEQALMTGYSDILYIRTCIPNNKPFMAEPPEV KQS" 642..2261 /gene="RSc0771" /note="RS05083" 642..2261 /gene="RSc0771" /EC_number="1.-.-.-" /function="small molecule metabolism" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (LARGE SUBUNIT) OXIDOREDUCTASE PROTEIN" /protein_id="CADI4473.1"

gene  
CDS  
/db\_xref="GI:17427783"  
/translation="MAETROAEQADITYVSGVAGALVATELARACKSVLMLAQRPL  
PRMELVEFFRIONAKMDMADYPTSPMAPIHEGPDNNYLLIGEEHFDQYIIRAVG  
TTHWAASTWRELPNDEKLRVYGIADWDLPLOYDLEERYGAEAAAGWGPNDDELG  
SPRSQYMAEPLPSFNERVYKDALNDMLPRLVTEPPVANSPPYGRPTCCGNNG  
MPCIGMAYNGLIYVEKAEQAGARLLENAAVYLEVGACKRIYAAHYKPKQYDRM  
EGKMNVLANKGIEPTKMLKMTSGADPFKQCNSEBDQNMHOPGCVGFYVDRLY  
FGKROEHTSLIGRDPDFRATQAGKKLHLSNIRIODETTRTFKAKKLKPAELDAR  
IRDOAARVQGDSPHEILPLEPNNIYPSAQTODLGIIPREIYRIDDYKSAVHR  
EYAAARVQGDSPHEILPLEPNNIYPSAQTODLGIIPREIYRIDDYKSAVHR  
SAPMTGVTVNTLIALALRIADQLEKKA"  
2263. 3552  
/gene="RSc0772"  
/note="RS05082"  
2263. 3552  
/gene="RSc0772"  
/EC\_number="1.-.-.-"  
/function="small molecule metabolism; energy transfer;  
electron transport"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=1  
/evidence=not\_experimental  
/product="PUTATIVE OXIDOREDUCTASE DEHYDROGENASE  
(CYTOCHROME C SUBUNIT) SIGNAL PEPTIDE PROTEIN"  
/protein\_id="CAD14474.1"  
/db\_xref="GI:17427784"  
/translation="MRKAKASFLLSVLGAAARADAPADAAQIKRGEYLAIACD  
CVACTHAGGKRFPAAGLPLOIPMLGTIYSSNITPDPYTGIGKMYEDEFBEARVQVSD  
DSENYTPAMPYPSYAKINDADMRLYAFERYGAAYOYAPPGTIIRPMLNMRPGLKLM  
NIALFKRNPYLPKGMQPTMNRGAVLNGAGLHGTCHTGPREFGAEMKAMERKGGYLA  
GSTLAGNSAVNTSPDPSGIGKMKPEQVOYLPRGSGYPTGQAGAPGCAVQNSFSM  
TERDRLAETRLTPAVSTGAQRARHGWGPDAVDIALRKPLLETTIDARLYLQNC  
ACHQADAPGPDGYPPPLHNSAVGADGTNLIQVMHGIERRAGDRHIGMPAFGRQ  
LSDAQALANVTYRFQCFDPPATPALTADEIKRMPQ"  
complement(3559. .5781)  
/gene="RSc0773"  
/note="RS05081"  
complement(3559. .5781)  
/gene="RSc0773"  
/note="RS05081"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=1  
/evidence=not\_experimental  
/product="PROBABLE TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD14475.1"  
/db\_xref="GI:17427785"  
/translation="MRTSARVILGIGTPIPIYAAATVEVLTEPDMNRKAPVLYNDHVSQ  
ANGREPLINGDALTMKKPEGESGRVAVPAPRLIASIDAIAGNDAROKPHAVTOLR  
TPVEALDLARIVYIPSVTLDTPLMILERSONRNMWTFEGADOSPDMKLTLEAE  
IATSGSVQDLDEVKRHLAATLTDVNOALITYASGTAIGSPAPPAKSAAGAASTRA  
PAAPDQDPYGIAMTVKGTNNAAIRSGKAGVYLRIQDARPPPIQADVYVGTTRID  
LAGLITSSLSUTLALHLGSAAMAHLYPLTGVVLVDLPDFDRGLIELRKGSSS  
WAEYDFTRVSGSDLGTLTAPMREPPEPQRLGELVSHOLFADLAPLIGASNS  
KORRAAPROAPADKLPVEPRTRWNAIDVDFTEGRIYRTADLIDHLYITKIO  
DAVLTLDPLNRYAGGTTISNLRDQSAAPQASALAAARKEIRQLPFNIESHAST  
GAINDARLSAGNSVALDASNGELITVEGTSKFIIEAGLVNVTITKLEGG  
DKVYNSINAAAGFVSDGLAQARTFVDTODAVIDYAGTSFKHESLDTIHPDPSKL  
RFYSRTPLUYAGTICHDVSVNPAVVALPAGGAVNALAFAPVAAYVPELGPAPDS  
PCGKLADVROGPTAPPPGVYRSQGTGRADGSPKDPAAAPARQALPLRDPPTAGG"  
5683. .6785  
/gene="RSc0774"  
/note="RS05080"

CDS		5883..6785 /gene="RS00774" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : probable Gene name confidence : hypothetical predicted by Homology predicted by Codon usage predicted by Framed"
gene		/transl_start=1 /transl_table=1 /evidence="not_experimental /product="PROBABLE TRANSMEMBRANE PROTEIN" /protein_id="CADI4476.1" /db_xref="GI:17427786" /translation="MSHVAISAPAHPPALDDVYARRRGVARLYGPALIEFFAAHIC VIGAGGWSWAELARCIGIRLLIDLDIAVSNTNRQJHALDGAKAIVDMAR IAEINPRAEISRIDDPVTVENVDALGHPEYVDAIDAIVAKTIVAFCQRTSRV TCGAGOLDPTRIKRVADLARTIODPLAKRGLRRHGFPDRPKAFGIDAVSDE PIKRPPEEOOACAVEPAESHAGIIDDLAAGAILEVNAPPAPPSNGPOLAACGGC SSCVTVTFVGKVASTALRAIAY" complement(6890..8257) /gene="kacsb" /note="RS00775; RS05079" complement(6890..8257) /gene="kacsb" /EC_number="1.11.1.6" /function="cell processes; protection responses; detoxification" /note="Product confidence : probable Gene name confidence : putative predicted by Homology predicted by Framed" /codon_start=1 /transl_table=1 /evidence="not_experimental /product="PROBABLE PEROXIDASE/CATALASE (CAT)" OXIDOREDUCTASE PROTEIN" /protein_id="CADI4477.1" /db_xref="GI:17427787" /translation="MGPEPEAAGLEEDLGWSSRFSGKGTDATISGLEVWTTPQC MSNPNLRCGYEWELTKSPAGAHOWAKDAETIPRADPSAKHLPMITLDSLRE VDNAEKTSRRYEHPPOFADFALAMRKUHLTRMDGPSARYLGEVPNALIMOPIRR VDHLLDTODIALKARILASLSVISOLVASPASSTFSRSDMRGGANGKITLAP KDWPANOPMAQLAKYLETESIQGFENGOSGKKVSLADLVLAGCAVEQANANAG AVEPEFTGRMDAQOTDYVESFAVLIEPADFPNYGKTYTLPAEALLDKADLLTI TAPMTYLVGTIRVIDNINIGOTRHGVFORRESLTDFNVLLDMGETMKATDGDVYA EGRRATGALKTKTIRVDIVGSHSQLALAEVYGSAQAOKKFVRDEFAAMDXYMNL
CDS		
Query Match	8.5%; Score 41; DB 1; Length 198050;	
Best Local Similarity	54.2%; Pred. No. 13;	
Matches	83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;	
QY	61 GCCCGACTTCAGCCGCCAGCAAGCTGGCCGCCGCACCGCGTGCGTGGCGGCTTT	120
Db 135919	GCTCCGGCCCAAAGCCCCGGATTCTTCGGCACCAATCATGGCGCGGTGGCGCTG	135978
QY	121 GCCGCCAGCTTGTGGTAAGGCTCTGGTCTGTGTGCATGGCCATCGCGCGGTT	180
Db 135979	CTGGGGCACACAGATCGGTGGCGGTACCGGCATACGGTGGCCACCATCGCGCGCGCC	136038
QY	181 CTGGGCGGTCTGATTTGGGCTTAATAATCGGTCAG	213
Db 136039	CTGGGTGCCCTGGCGGCAATCAGGTGAGCGG	136071
RESULT 8		
LOCUS	AL646079	213050 bp DNA linear BCT 07-DEC-2001
DEFINITION	Raisstonia solanacearum GM1100 megaplasmid, complete sequence;	
ACCESSION	AL646079 AL646053	
VERSION	AL646079.1 GI:17430956	



		/transl_table=11	/evidence=not_experimental	/product="PROBABLE COBALUT-ZINC-CADMIUM OUTER MEMBRANE RESISTANCE PROTEIN"	/protein_id="CADI7642.1"	/db_xref="GI:17430961"	/translation="NWTHPCIRAAVAALSCALLVRSGLAAPGEEAAMASDTAPAAOS APASVGAAHEMTLLOOAILDRALATHPALAAAGYAVRASGAVDQASRLRNPFAFLT EDVRBORATHTAOINTPLEIGGRKARTRAEFLARPAGDDAVLAELRADVTBAEF ELSIAOERLRILAAQSALAQAQAAAVRRORAKGSVPVCOTKAIVARACAIETLRRA ADLLVAQRSLAFWGGVAAYVEADSGSETTIPAPPRAHRYLAATAEYPYALASTPRA LARLEIEHRQALVNVERSKRPDVTLLYSGARSDSGANMAVVGIATPLPDRNQG SLIEQALRDKAEDYRALRIEQAAALSOPVALTEARAQAVGALRODVLPFGAGRADA ARIGFEGKGFNFIDVIDAORTLEQARAQYLAAISRAHOAAAGIDRTITGR"		
gene		5752..6966	/gene="czcb"	/note="RSP0492; RS00370"	/gene="czcb"	5752..6966	/function="cell processes; protection responses"	/note="Product confidence : probable Gene name confidence : probable predicted by Homology predicted by Framed" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="PROBABLE COBALUT-ZINC-CADMIUM RESISTANCE (CAUTION EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"	
CDS							/protein_id="CADI7643.1"	/db_xref="GI:17430962"	/translation="MPLPRQORRAIALIVALGALAGALMWGMGAROAHAASEADQP AAHAEHQAPKAEEAAVEAHMETIAMEATANAIGVETGAOLREVLPEELIRLN DDRTAHYPPRVAGAEVAGEVATELGOPYRGQLAVLSSPALSDORSIELQAQORLIAAO EAHAREKRIMWEOGIIAAODYOQARTALOEARIADANROKLIAGAIPAGTAHNREFL RAPDPGYVARRHLSOGSANOEAFAVFPIADLTRYMADFAYTAIDLAEVTHATAYTRA TATGTAVQGVSTVYGALLGEOTRSAPARYTLDNPKRAMRGMYSVSESGRYPVPA VRADAVGTDEKPSVEFVPPAPGFEVPRVKGRSDGTETLLDGIPAGTRYAANSTVL KAEKGKSADHG"
gene		6977..10171	/gene="czca"	/note="RSP0493; RS00371"	/gene="czca"	6977..10171	/function="cell processes; transduction of small molecules"	/note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage	
CDS									
Query Match	8.4%	Score 40.8:	DB 1;	Length 213050;					
Best Local Similarity	58.1%	Pred. NO.15;	Mismatches 52;	Indels 0;	Gaps 0;				
Matches	72;	Conservative 0;							
Oy	85	GTTCGCCCGCCACCCTGCCTGTTGGCGCGGTGTTGCCCGCCACTTGCGTAAAGGC	144						
Dd	20248	GTGGGCGCGGCTCTGTGGCGGCGGTGCTGCGGCGGCTGTCGACACCAAGATCGTGGCGGG	20307						
Oy	145	TCTGCTCGTGTGTCATGCCCATCGCGCGTCTGGCGGCTGTGCGGCTGTGATGGCTCTTAA	204						
Dd	20308	CGCGGCAATACGGTGGCCCATCGCGCGGCTGTGGCGGCGCGCTGCGCGGCAACACC	20367						
Oy	205	ATCG 208							
Dd	20368	ATCG 20371							
RESULT 9	RIBANT17KA	539 bp	DNA	linear	BCT 26-APR-1993				
LOCUS	R.conor11	17k	genus-common antigen gene,	complete cds.					
DEFINITION	M28480								
ACCESSION	M28480.1	GI:152457							
VERSION									

KEYWORDS	antigen.
SOURCE	R. conorii DNA.
ORGANISM	Rickettsia conorii
REFERENCE	Bacteria: Proteobacteria: alpha subdivision; Rickettsiales: Rickettsiaceae: Rickettsiae; Rickettsia: spotted fever group.
AUTHORS	Anderson, B.E. and Tzianabos, T.
TITLE	1 (bases 1 to 539)
	Comparative sequence analysis of a genus-common rickettsial antigen gene
JOURNAL	J. Bacteriol. 171, 5199-5201 (1989)
MEDLINE	89359171
COMMENT	Draft entry and printed copy of sequence for [1] kindly submitted by B. Anderson, 27-OCT-1989.
FEATURES	Location/Qualifiers
source	1..539
	/organism="Rickettsia conorii"
-35_signal	/db_xref="taxon:781"
-10_signal	1..6
mrna	21..27
	38..539
	/product="17 kd mrna"
CDS	60..539
	/note="17 kd antigen precursor"
	/codon_start=1
	/transl_table=1
	/protein_id="AA26376.1"
	/db_xref="GI:152458"
	/translation="MKLLSKIMTIALATSMIQACNCPGKNNKQGTGLLGAGGALLG SOFKCGQIVGVGVALGAVLGGQIGAGNDEDRRLAELTSORALETAPSGSNEM RNPNGNRYGVYTPPKRTYRNSYGYCREYTGTVVYGGKRAYGNACRPDGVQVNA"
sig_peptide	60..119
	/note="17 kd antigen signal peptide"
mat_peptide	120..536
	/product="17 kd antigen"
BASE COUNT	179 a 125 g 142 t
ORIGIN	
Query Match	8.3%; Score 40.2; DB 1; Length 539;
Best Local Similarity	46.1%; Pred. No. 18;
Matches 172; Conservative	0; Mismatches 198; Indels 3; Gaps 1;
QY	100 GGTGGGTTGTGGGCGGTGTGCGGCCAGCTGTCGTAAGAAGCTCTGCTGTGTCG 159
DB	165 GCGGTGCTGGCGCGCAATTACTGTCTCAATTCGGTAAGCGCAAGACGCTTGT 224
QY	160 ATGGCCATCGGGGCGGCTGTGGCGGTCGTGATTGGCTTAATATCGATCGACAGCATG 219
DB	225 GGAGTAGGTGTAGGTGACATTACTTTGAGACAGTCTTGTGTGGACAAATCGTCCAGGTATG 284
QY	220 GACCAAGATTAATCAACTGAACCAAGTCCTGTGAAAAAGTAAACCGCGCAGGTT 279
DB	285 GATGAACAGAGTAGA--AGACTTCAGAGGCTTACTCTACAGAGAGCTTTAGAAACACT 341
QY	280 ACTGCTGGCGTAATCCGACACCGGTAAACACTCTACTGTGGAACCGGTTGACACTTAC 339
DB	342 CCTAGGTGTAGTAGCTAATATGGCTAATCCGGTAACGCAATTTAGCGTTACTATACA 401
QY	340 CAGCGTTACAAACAAACAGGACCGCGTCAGCATCTACTCGCGGAAATTCAGCAGAAAGCC 399
DB	402 CCTAATAAAACTTATATGAATATAGCACTGGTCAATATTGGCGTAGACACTCAACAGTT 461
QY	400 ATGATCGCAGGTTCAGAACAGAAATCTACGGCACCCGCTGGCCCTCAGCGGATGGCGCG 459
DB	462 GTAATATGGCGGAAACACAAAGAGCATCGGTATATGATCATGTCGCCCACTGACGAGACA 521
QY	460 TGGCAGGTGATTA 472
DB	522 TGGCAGGTGTGCA 534
RESULT 10	
RIRANT17KD	
LOCUS	RIRANT17KD 539 bp DNA linear BCT 26-APR-1993



DEFINITION R.rickettsia 17k genus-common antigen gene, complete cds.  
 ACCESSION M28479  
 VERSION M28479.1 GI:152463  
 KEYWORDS antigen.  
 SOURCE R.rickettsia DNA.  
 ORGANISM Rickettsia rickettsii  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Rickettsiaceae; Rickettsiinae; Rickettsia; spotted fever group.  
 REFERENCE 1 (bases 1 to 539)  
 AUTHORS Anderson, B.E. and Tzianabos, T.  
 TITLE Comparative sequence analysis of a genus-common rickettsial antigen gene  
 JOURNAL J. Bacteriol. 171, 5199-5201 (1989)  
 MEDLINE 89359171  
 COMMENT Draft entry and printed copy of sequence for [1] kindly submitted by B.Anderson, 27-Oct-1989.  
 FEATURES  
 source  
 1. .539  
 Location/Qualifiers  
 /organism="Rickettsia rickettsii"  
 /db\_xref="taxon:783"  
 1. .6  
 21. .27  
 37. .>539  
 /product="17 kD mRNA"  
 60. .539  
 /note="17 kD antigen precursor"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="AA26379.1"  
 /db\_xref="GI:152464"  
 /translation="MKLSKIMIALATSMLOACNPGMKNKOGTTLIGAGALLG  
 SOFGKQGLGVGALLGAVLGIGGIDGMDQBRRLAELTQSRLALEAPGSSNVEW  
 RMPDNGNYGYTPNKTYYRSTQYCRRETYQYVIGGKQKATGNACRDPDGMQVNV"  
 60. .119  
 /note="17 kD antigen signal peptide"  
 mat\_peptide 120. .536  
 /product="17 kD antigen"  
 BASE COUNT 179 a 92 c 125 g 143 t  
 ORIGIN  
 Query Match 8.3%; Score 40.2; DB 1; Length 539;  
 Best Local Similarity 46.1%; Pred. No. 18;  
 Matches 172; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

RESULT 11  
 LOCUS AE008675  
 DEFINITION Rickettsia conorii Malish 7, section 107 of 114 of the complete genome.  
 ACCESSION AE008675 AE006914  
 VERSION AE008675.1 GI:15620413  
 KEYWORDS  
 SOURCE Rickettsia conorii.  
 ORGANISM Rickettsia conorii  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Rickettsiaceae; Rickettsiinae; Rickettsia; spotted fever group.  
 REFERENCE 1 (sites)  
 Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E., Raoult, D. and Claverie, J. M.  
 Selfish DNA in protein-coding genes of Rickettsia  
 Science. 290 (5490), 347-350 (2000)  
 JOURNAL MEDLINE 20485642  
 PUBMED 11030655  
 REFERENCE 2 (sites)  
 Ogata, H., Audic, S. and Claverie, J.-M.  
 Selfish DNA and the origin of genes  
 Science 291 (5502), 252-253 (2001)  
 JOURNAL 3 (bases 1 to 10127)  
 REFERENCE Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.  
 Mechanisms of Evolution in Rickettsia conorii and R. prowazekii  
 Science. 293 (5537), 2093-2098 (2001)  
 JOURNAL 4 (bases 1 to 10127)  
 REFERENCE Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.  
 Direct Submission  
 Submitted (26-Apr-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France  
 A public version of R. conorii genome database is accessible at http://igs-server.cnr-mrs.fr/. The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.  
 FEATURES  
 Location/Qualifiers  
 1. .10127  
 /organism="Rickettsia conorii"  
 /strain="Malish 7"  
 /db\_xref="taxon:781"  
 111. .890  
 /gene="map"  
 /note="RC1276"  
 111. .890  
 /gene="map"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methionine aminopeptidase [EC:3.4.11.18]"  
 /protein\_id="AA03814.1"  
 /db\_xref="GI:15620414"  
 /translation="MTIKIHGKDFIKRMAAGKLAETLDTTHDKPNVTNSLGL  
 CNFTTSHNAPAPLNKGEFKSICSTINHHVCHGIPNDKPLKNGDIVADVTLLDG  
 WYGDTSRMVYGVDAIKPKRLIOVYTDMMKGIIEVVRGAKGLDIGHAIOSYAEKHY  
 SVYRDYGHGIRVRHDEPSILNYSRGTGLTEGMPFVPPMHNACNYDTITLSKLD  
 GMYVTRDRLSLAOGFENHTIGVTKDFEFTTSPKKDVPY"  
 complement(977. .10111)  
 /note="REP03, repeated element"  
 complement(1064. .1083)  
 /note="REP02, RS3-like repeat"  
 1141. .1452  
 /gene="RC1277"  
 1141. .1452  
 /gene="RC1277"  
 /codon\_start=1  
 /transl\_table=11  
 repeat\_region  
 repeat\_region  
 gene  
 CDS

```

/product="similarity to DNA repair protein (RadC)"
/protein_id="AA03815.1"
/db_xref="GI:15620415"
/translation="MLENNLEFSNDEIDIPHYIGHKRVRKRPVAGAYESDYEL
LEVMLSALIPKRIKPLAKRLDHPDIDLI.NLDERLSTIKGTENENLYINFAIRELI
N"
1457. .1825
/gene="RC1278"
1457. .1825
/gene="RC1278"
/codon_start=1
/transl_table=11
/product="similarity to DNA repair protein (RadC)"
/protein_id="AA03816.1"
/db_xref="GI:15620416"
/translation="MKRIKKNNINASSMWDYKYNMGNNRLEQFRILFLNKNIL
IADEVLSQGTIDQAAVYPREIIRKALRNENASLILVHHHPSGSPESKANIMHTKIV
BRCOFINIVNLCSYEYK"
1812. .2795
/gene="maray2"
1812. .2795
/gene="RC1279"
1812. .2795
/gene="maray2"
/codon_start=1
/transl_table=11
/product="undecaprenyl-phosphate
alpha-N-acetylglucosaminyltransferase [EC:2.4.1.-]"
/db_xref="GI:15620417"
/translation="MNINSYTLFIFFSFISPAILFYILTKYLIPTIGLVDPISNRSD
KITRGGAIAIVFVMAISGFYITSNFNVSIRKIPILAIASISFLDMLAAPTLL
LRILVHICAFALFELFQSLNHSVHILYISIVIASGFTINYNMDSIDMLCAVPLL
LSSTMLICFLOPPTINPNPFTSYVNVITLIGSGFLEFMNPAKIFLGVGVSISGF
LNLGILLALATSTFACATASLYITDAVLTLLRLNKEKIMQPHLKHFFQNAV
2792. .3115
/gene="RC1280"
2792. .3115
/gene="RC1280"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA03818.1"
/db_xref="GI:15620418"
/translation="MKLIKEDIDKIVHHIFAKOHPPLPEIMINMKNKIVGFNPSKALP
LKITYTYKKOKINTLFTIQADENATAELPYODILIERIKIYLGFAIHQMNVTFTK
GKKSIL"
complement(3198. .3227)
/feature="REP02, RS3-like repeat"
3113. .3353
/feature="REP01, repeated element"
complement(3509. .4258)
/gene="RC1281"
complement(3509. .4258)
/gene="RC1281"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA03819.1"
/db_xref="GI:15620419"
/translation="MKKILLIAAASSTALLSGLSFADCDNMSSVSSTSSMSSVEN
OMYIKINAGVGFENKTPKRGADFKLNNKSNKISNGTGTGTEIGAGYIMDNLTDLTI
GTVASHLKSKSKTTPDGSFSVKNKPTIVSYLNGYDFPDLIMFVYFAGAGYGAFAV
KKKHSKDIKGGVNDTENGTTKNKTNFAVSLSGTSEVAGQVKAELVYSMDYGRKTK
NTKTIINDGVKFGFTGTHYKGNHLMAGLRPFM"
4438. .4463
/feature="REP03, repeated element"
complement(4505. .5188)
/gene="RC1282"
complement(4505. .5188)
/gene="RC1282"
/codon_start=1
/transl_table=11

```

```

/product="unknown"
/protein_id="AA03820.1"
/db_xref="GI:15620420"
/translation="MLIMKRLILIAATSATMLSSSYFAEGMDHEMYLRIDGAMFN
EEKDARGVKLRKSNWTVPVALGIGYISSENFADLTGTTGGKLRKSGAATNAFTR
TNISASHKPTITLLINGYDLGNFMPFVAGAGYSAVLEKITYNGITGLSNTK
NRNINYSKLTIGTSAQIADGVKELVYSWIDGRKSKNVIYQSTVPTGMYQSHN
LTAGIRFDI"
complement(5338. .5508)
/feature="REP06, A and T rich repeated element"
complement(5478. .5828)
/gene="fdaX"
/feature="RC1283"
complement(5478. .5828)
/gene="fdaX"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="AA03821.1"
/db_xref="GI:15620421"
/translation="MTYVVDCEVKCKYTDCEVPCPYDCEYEGEERDDFLVINPDEC
IDCGVCPDPCPIGAIKPESPLIEWERANDFIENKGMKNITKKTALPDADKFDKE
DKFNKYGISVKIAN"
complement(5833. .6525)
/gene="csmC"
/feature="RC1284"
complement(5833. .6525)
/gene="csmC"
/codon_start=1
/transl_table=11
/product="heme exporter protein C"
/protein_id="AA03822.1"
/db_xref="GI:15620422"
/translation="MKLPIFYENKLSVILPILASTCVMMITGLTIALIVSPDY
OGDPEVRIWVHHVPSAMMALGVYVEMAVCSFSLVWKTITISYLLAVASGATFTIL
SIGLNTGKGRPIWGWVMDARLTSMILFLYLSYITIVNSADINIRKQAPRSITIL
RTAILIDIKIKNR"
complement(6540. .7046)
/gene="RC1285"
complement(6540. .7046)
/gene="RC1285"
/codon_start=1
/transl_table=11
/product="cytosine deaminase [EC:3.5.4.1]"
/protein_id="AA03823.1"
/db_xref="GI:15620423"
/translation="MNILDKKANPNFMDALKQAKIAFDKNEVPVGAVVVDRLHQ
KIATSHNNTTEKNALYHAEIINAEKDISKNLNDYDLYLLEPCAMCAAAIAH
SLKRLIFYGASDSKHGVESNLRKFNSSACFHRPEIYSIGLAEBSGLMKKEFFKRIPT
VSSHRRMT"
complement(7027. .7947)
/gene="p34"
/feature="RC1286"
complement(7027. .7947)
/gene="p34"

```

Query Match 8.3%; Score 40.2; DB 1; Length 10127;  
Best Local Similarity 46.1%; Pred. No. 19;  
Matches 172; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

```

Db 100 GGTGCGGTGTGGCGGTGTCGCGGACCTGTGTAAGACCTCTGCGTGTGTCG 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8239 GCGGTGCTGGCGGCGCATCTACTGTTCAATTCGTAAGGCAAGACGCTGT 8298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 ATGCCCATCGCGGTCGGTTCGGGCGGTCGATTCGCTTAATCGTCAGAGCATG 219
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8299 GCACTAGTGTAGGTGATCTTGTGACGACAGTCTTGTGACCAATGCGTCAGTATG 8358
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 GACCAGCAGATTAATCAACTGAACAGTCTGTGAAAAAGTGAAGCGCGGACGTT 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8359 GATGACAGGATAGA--AGACTTGCAGAGCTTACCTCAGAGAGCGCTTTAGAAACAGCT 8415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



Query Match	8.24;	Score 39.6;	DB 1;	Length 13946;
Best Local Similarity	56.0%;	Pred. No. 27;		
Matches 75;	Conservative	0;	Mismatches 59;	Indels 0;
			Gaps	
Qy	86	TTGGCGGCGGCACCGGTTGTGGGCGGTTCGCCGTCAGCTGTTCGTAAGGCT	145	
Db	7494	TCGGTTCATCGCGCGGCGCGGTCTCCGCGAGCGCGGCGGCGCA	7533	
Qy	146	CTGTGCTGTGTGATGCCATCGCGGCGGTTCGGGCGGTGTGATGGCTCTAAAA	205	
Db	7554	AGGCGACCTACGTCCGCGCATATCGCGCGGTCTCGCTCGCGCTCTGAGGCGCTA	7613	
Qy	206	TCGGTCAGAGCATG	219	
Db	7614	CCGAGAGAGGCGCTG	7627	
RESULT 13				
ATF7K2				
LOCUS		106702 bp	DNA	linear
DEFINITION	ATF7K2			PLN 10-MAR-2000
	Arabidopsis thaliana DNA chromosome 4, BAC clone F7K2 (ESSA			
ACCESSION	AL0033545			
VERSION	AL0033545.2			
KEYWORDS	GI:5738368			
SOURCE				
ORGANISM	thale cress			
	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi			
REFERENCE	Bevan, M., Medler, H., Wambutt, R., Bancroft, I., Mewes, H.W., Lemcke, K.			
AUTHORS	and Mayer, K.F.X.			
	Unpublished			
JOURNAL	2 (bases 92830 to 92989)			
REFERENCE	Bevan, M., Volckaert, G., Gymnopoulos, B., Voet, M., Robben, J.,			
AUTHORS	Bancroft, I., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 16021)			
AUTHORS	Bevan, M., Robben, J., Gymnopoulos, B., Volckaert, G., Bancroft, I.,			
	Mewes, H.W., Lemcke, K. and Mayer, K.F.X.			
	Unpublished			
JOURNAL	4 (bases 1 to 106702)			
REFERENCE	EU Arabidopsis sequencing project.			
AUTHORS	Submitted (09-MAR-2000) MIPS, at the Max-Planck-Institut fuer			
TITLE	Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG, E-mail:			
JOURNAL	lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project			
	Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge			
	Laboratory, John Innes Centre, Colney Lane, NR4 7JF Norwich, UK,			
	E-mail: michael.bevan@jnc.ac.uk			
COMMENT	On Aug 18, 1999 this sequence version replaced gi:3892698.			
	Information on performance of analysis and a more detailed			
	annotation of this entry and other sequences of chromosomes 3, 4			
	and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.			
	Location/Qualifiers			
	1..106702			
FEATURES				
Source	/organism="Arabidopsis thaliana"			
	/variety="Columbia"			
	/db_xref="taxon:3702"			
	/chromosome="4"			
	1..15769			
misc-feature	/note="basepairs 1-15769 are not integral part of BAC			
	clone F7K2			
	complement(join<3..98,594..821))			
CDS	/gene="AT4G22380"			
	/note="strong similarity to NHP2/R56 FAMILY PROTEIN, Homo			
	sapiens, PID:g4826860			
	contains Ribosomal protein L7Ae signature AA73-90			
	contains EST gb:f13981"			
	/codon_start=1			
	/product="Ribosomal protein L7Ae-like (fragment)"			

exon /protein\_id="CAB52812.1"  
/db\_xref="GI:5738369"  
/translation="MTGEVNPVKAYPLADSOIITIMDI.VQATNYKOLKKGANEATK  
TUNRGISFEVWMAADAEPLLEILHLPLADEKNVPIVFPVKALGRACGAVRPVIAC  
SVTSNE"  
complement(3..98)  
/gene="At4g22380"  
/number=1  
3..821  
/gene="At4g22380"  
complement(join(3..98,594..821))  
/gene="At4g22380"  
complement(99..593)  
/gene="At4g22380"  
/number=1  
complement(594..821)  
/gene="At4g22380"  
/number=2  
1789..2973  
/gene="At4g22390"  
complement(1789..2973)  
/gene="At4g22390"  
complement(1789..2973)  
/gene="At4g22390"  
/number=1  
complement(1789..2973)  
/gene="At4g22390"  
/note="similarity to various predicted proteins,  
Arabidopsis thaliana"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB52813.1"  
/db\_xref="GI:5738370"  
/translation="MAECPTDLINEMFLRLRATTLVKCVLSKPCSLIDSEPNVSSH  
LRRLETGELHMLLRGPRLRVLDESSENVSDIPHPLOAGGTEVFSGFNGVIGLC  
NSPVDLAIFNPSTRIKIHRLPIEDFPERDITREYVFGIDGVSDDDKVVAIVQCK  
LKEGKKFPCPEVKVFSLKNSMKRVCAMFEFQILMIYYHLLPRRGVYVNNHL  
HHILPRGCVIAFNATIKYDLASDDIGVSPFOELIENMDIGVLGVCVLCMYDEY  
SHVDVWVLRKEVEDYKSMTKLYVRPKESVESVFETPLICSDRSKILLEINNAALM  
WFDLESQSLTITAGIECDSSFTADILVSSLVLGCGDPTQAQNSKDKMKPKSTKRMYA  
LL"  
4728..5711  
/gene="At4g22400"  
/number=1  
4728..5711  
/gene="At4g22400"  
4728..5711  
/gene="At4g22400"  
/note="similarity to various predicted proteins,  
Arabidopsis thaliana  
contains EST gb:217595"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB52814.1"  
/db\_xref="GI:5738371"  
/translation="MSSHLISQICSCRDPLGNLPADLIRKCTDLMDLMPAGSORLRT  
LKNKKTALPSFRKAYRREPRMLLYREGSGETRFDPVYRREYVHGNPNRLADAEFL  
GSTLGLVMSLIDLNPRTHQFTFLNPTISELQQLPELTLDSPTGLVLRGVLTLGN  
PSDNNTYACLITDYLHETIGNRRDTYLLLYVAKTSGRWEQNMVSASMLVYGIFDTWY  
IESISPSDKISIEIWPKRYDFTFOENNVSNFTIPVDPIOMDPFDOSLYHKHRLQ  
LPEPSTLAITIGTSPDPRVRNKKIIGQSDDDPRASDHIIIDGVVVEVRP"  
6112..8034  
/gene="At4g22410"  
complement(join(6112..6297,6675..6812,6908..7075,  
7168..7353,7429..7513,7775..8034))  
/gene="At4g22410"  
complement(6112..6297)  
/gene="At4g22410"  
/number=1  
complement(join(6112..6297,6675..6812,6908..7075,  
7168..7353,7429..7513,7775..8034))  
/note="similarity to snRNP assembly factor like protein,"

Homo sapiens, PID:g5730025"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB52815.1"  
/db\_xref="GI:5738372"  
/translation="MVEFQVLDFHFERPCSVSLSNLNVYACLVGKRYFGGRSQSHAY  
TSLSEAGHHVYINLLEKYYCLDPSYIENDPSLDIRHLNPFSSRAQVNELDKNQW  
SRALQSDYLPGWGLNNTIOKTFEVSITIOSLMRVRPLNPNFPIPENYOHCKSPLVHC  
FGELETRKTIWHRARFKGOVSPHEIQAVMKASKRPRIQGOOPVEFMSLNTLHNDL  
RTSKDASITIHKCFQGELEVKEFGQENKEIISKMSFLMLGLDLPPLPKDVKENKI  
IPQVALFDLLKFDGSEVTEVAPKILARRKRYIKSPRYLMPHVRKKNPFKEKNP  
TLGESV"  
complement(6298..6674)  
/gene="At4g22410"  
/number=1  
complement(6675..6812)  
/gene="At4g22410"  
/number=2  
complement(6813..6907)  
/gene="At4g22410"  
/number=2  
complement(6908..7075)  
/gene="At4g22410"  
/number=3  
complement(7076..7167)  
/gene="At4g22410"  
/number=3  
complement(7168..7353)  
/gene="At4g22410"  
/number=4  
complement(7354..7428)  
/gene="At4g22410"  
/number=4  
complement(7429..7513)  
/gene="At4g22410"  
/number=5  
complement(7514..7774)  
/gene="At4g22410"  
/number=5  
complement(7775..8034)  
/gene="At4g22410"  
/number=6  
8059..8449  
/note="389bp LTR"  
8450..13234  
/note="LTR retrotransposon"  
13234..13624  
/note="389bp LTR"  
complement(13513..13618)  
/gene="At4g22420"  
/number=1  
complement(join(13513..13618,13701..14033,14216..14331))  
/gene="At4g22420"  
13513..14331  
/gene="At4g22420"  
complement(join(13513..13618,13701..14033,14216..14331))  
/note="similarity to predicted protein, Arabidopsis  
thaliana  
contains EST gb:T44427"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB52816.1"  
/db\_xref="GI:5738373"  
/translation="MAECPTDLINEMFLRLRATTLKCRVLSKPCSLIDSEPKRYIE  
RSDSPPLPGFNHLLPLANAYDDDEEGNELKSOARRNGVAKGECGNKNYVGEAO  
REVDEEDDDDDSKGRKSHRVEVRDPCPYLDVTNPOVITIIDPLMLRVPLATMRK  
RMRTGGRKAKAMEKYLKXVKKSG"  
complement(13619..13700)  
/gene="At4g22420"  
/number=1  
complement(13701..14033)  
/gene="At4g22420"

Matches	Accuracy	Pred. No.	Length
67; Conservative	59.38;	31;	106702;

	0; mismatches	46;	Indels	0;	Gaps	0;
85 GTTGGCGGCGCACCGTGCGGTGTGGCGGCTCCGCCGCTC						

**C**

33426 GTTGGAGGTGTCCTCCCGTGTTGATGTTTTGCACCAATTGCTC  
||||| ||||| |  
.....GTTGCCCGCCACACTGTTCCGTAAGAACC 144

145 TCTGGTCGTGCTCGATGCCATCCGCCGTGCAGTTCCTGGAAAGTGTC 33485  
||||| ||||| | ||||| ||  
33486 TCTGGTCGTGCTTTGGAGGACCTCCCCC ||||| |||||

-----GCGGGTGTGTGGAGCTTTCTGC 33538

ATRC157  
 Arabidopsis thaliana DNA  
 195975 bp  
 contig fragment No. 57.

ORGANISM  
thale cress.  
*Arabidopsis thaliana*  
Eubacterium

Eukaryota; Viridiplantae; Streptophyta; Embryophyta. Archaeplastida; Rhodophyta; Charophyta; Magnoliophyta; Spermatophyta; Magnoliophyta.

[illegible]

Medler, H., Wambutt, R., Mewes, H.W., Lemcke, K and Wambutt, R. (1993) Unpublished

2 (bases 141638 to 141797)  
Volckaert, G. and Mayer, K.F.X.

Lemcke, K. and Mayer, K.F.X.  
 Voet, M., Robben, J., Mewes, H.W.

3 (bases 143670 to 199577)  
Hilbert "

Waidel, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft A., Mewes, H.-W., Lemcke, K. and Mayer v. D. (1996).

4 (bases 48809 to 64829)  
unpublished  
Mayer, A.F.F.X.

Volckaert, G., Mewes, H W  
Mayer, K.F.X.

Unpublished  
5 (bases 1 to 199577,  
5

EU Arabidopsis sequencing project.  
Direct Submission

Submitted (10-MAR-2000) MIPS, at the May-1999 Biochemie

lemcke@ips.biochem.mpg.de, mayer@mps.biochem.mpg.de

Correspondence: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney, Cambridge CB3 0ET, UK.  
E-mail: [mike.bevan@jic.ac.uk](mailto:mike.bevan@jic.ac.uk)  
Fax: +44 (0)1223 336000

Information on performance of annotated

annotation of this entry and a more detailed analysis and other sequences of chromosome 3 and 5 can be viewed at: <http://4-1000genomes.org>

This fragment has an overlap with ATCHRIY56 at the 5' end overlap with ATCHRIY58 at the 3' end.

Location/Qualifiers  
1. 199577

```

/organism="Arabidopsis thaliana"
/variant="Col-0"

```

```

    /db_xref="taxon:3702"
    /chromosome="1"

```

```

?cut.unosome="4"
6674. .9057
/row= "1.1"

```

```

/ gene="AT4g22260"
complement(6674..6882)

```

```
/gene="AT4g22260"  
/number=1
```

```
complement(join(6674. .6882,7268. .7339,7507
7667. .7775,7863 7970 8077 8175 8272 8370 8467 8564 8661 8758 8855 8952 9049 9146 9243 9340 9437 9534 9631 9728 9825 9922 10019 10116 10213 10310 10407 10504 10601 10698 10795 10892 10989 11086 11183 11280 11377 11474 11571 11668 11765 11862 11959 12056 12153 12250 12347 12444 12541 12638 12735 12832 12929 13026 13123 13220 13317 13414 13511 13608 13705 13802 13899 13996 14093 14190 14287 14384 14481 14578 14675 14772 14869 14966 15063 15160 15257 15354 15451 15548 15645 15742 15839 15936 16033 16130 16227 16324 16421 16518 16615 16712 16809 16906 17003 17100 17197 17294 17391 17488 17585 17682 17779 17876 17973 18070 18167 18264 18361 18458 18555 18652 18749 18846 18943 19040 19137 19234 19331 19428 19525 19622 19719 19816 19913 20010 20107 20204 20301 20398 20495 20592 20689 20786 20883 20980 21077 21174 21271 21368 21465 21562 21659 21756 21853 21950 22047 22144 22241 22338 22435 22532 22629 22726 22823 22920 23017 23114 23211 23308 23405 23502 23599 23696 23793 23890 23987 24084 24181 24278 24375 24472 24569 24666 24763 24860 24957 25054 25151 25248 25345 25442 25539 25636 25733 25830 25927 26024 26121 26218 26315 26412 26509 26606 26703 26800 26897 26994 27091 27188 27285 27382 27479 27576 27673 27770 27867 27964 28061 28158 28255 28352 28449 28546 28643 28740 28837 28934 29031 29128 29225 29322 29419 29516 29613 29710 29807 29904 30001 30098 30195 30292 30389 30486 30583 30680 30777 30874 30971 31068 31165 31262 31359 31456 31553 31650 31747 31844 31941 32038 32135 32232 32329 32426 32523 32620 32717 32814 32911 33008 33105 33202 33300 33396 33493 33590 33687 33784 33881 33978 34075 34172 34269 34366 34463 34560 34657 34754 34851 34948 35045 35142 35239 35336 35433 35530 35627 35724 35821 35918 36015 36112 36209 36306 36403 36500 36597 36694 36791 36888 36985 37082 37179 37276 37373 37470 37567 37664 37761 37858 37955 38052 38149 38246 38343 38440 38537 38634 38731 38828 38925 39022 39119 39216 39313 39410 39507 39604 39701 39798 39895 39992 40089 40186 40283 40380 40477 40574 40671 40768 40865 40962 41059 41156 41253 41350 41447 41544 41641 41738 41835 41932 42029 42126 42223 42320 42417 42514 42611 42708 42805 42902 43000 43096 43193 43290 43387 43484 43581 43678 43775 43872 43969 44066 44163 44260 44357 44454 44551 44648 44745 44842 44939 45036 45133 45230 45327 45424 45521 45618 45715 45812 45909 46006 46103 46200 46297 46394 46491 46588 46685 46782 46879 46976 47073 47170 47267 47364 47461 47558 47655 47752 47849 47946 48043 48140 48237 48334 48431 48528 48625 48722 48819 48916 49013 49110 49207 49304 49401 49498 49595 49692 49789 49886 49983 50080 50177 50274 50371 50468 50565 50662 50759 50856 50953 51050 51147 51244 51341 51438 51535 51632 51729 51826 51923 52020 52117 52214 52311 52408 52505 52602 52700 52796 52893 52990 53087 53184 53281 53378 53475 53572 53669 53766 53863 53960 54057 54154 54251 54348 54445 54542 54639 54736 54833 54930 55027 55124 55221 55318 55415 55512 55609 55706 55803 55900 56000 56096 56193 56290 56387 56484 56581 56678 56775 56872 56969 57066 57163 57260 57357 57454 57551 57648 57745 57842 57939 58036 58133 58230 58327 58424 58521 58618 58715 58812 58909 59006 59103 59200 59297 59394 59491 59588 59685 59782 59879 59976 60073 60170 60267 60364 60461 60558 60655 60752 60849 60946 61043 61140 61237 61334 61431 61528 61625 61722 61819 61916 62013 62110 62207 62304 62401 62500 62596 62693 62790 62887 62984 63081 63178 63275 63372 63469 63566 63663 63760 63857 63954 64051 64148 64245 64342 64439 64536 64633 64730 64827 64924 65021 65118 65215 65312 65409 65506 65603 65700 65797 65894 65991 66088 66185 66282 66379 66476 66573 66670 66767 66864 66961 67058 67155 67252 67349 67446 67543 67640 67737 67834 67931 68028 68125 68222 68319 68416 68513 68610 68707 68804 68901 69000 69096 69193 69290 69387 69484 69581 69678 69775 69872 69969 70066 70163 70260 70357 70454 70551 70648 70745 70842 70939 71036 71133 71230 71327 71424 71521 71618 71715 71812 71909 72006 72103 72200 72297 72394 72491 72588 72685 72782 72879 72976 73073 73170 73267 73364 73461 73558 73655 
```

```

/feature="AT4g22260". .8454,8588. .8777,8909.
complement(join(c574

```

7667. .7775,7863. .7978,8361. .8454,8588  
/gene="ATM"0014. .6882,7268. .7339,7507. .7575,

6068'///8: .0000/7-2  
"007776377

Intron	/gene="AT4g22270" /number=1 10395..10516 /gene="AT4g22270"
exon	/gene="AT4g22270" /number=2 10517..11326 /gene="AT4g22270"
Intron	/number=2 11168..11253 /gene="AT4g22270"
tRNA	/note="Tyr-tRNA <sub>(GTA)</sub> " /product="tRNA-Tyr" 11327..11779 /gene="AT4g22270"
exon	/number=3 11780..11867 /gene="AT4g22270"
Intron	/number=3 11868..12174 /gene="AT4g22270"
exon	/number=4 12175..12253 /gene="AT4g22270"
Intron	/number=4 12254..12393 /gene="AT4g22270"
exon	/number=5 12761..15111 /gene="AT4g22280"
gene	complement(12761..13015,13081..13227,13379..14157, 14209..14314,14995..15111) /gene="AT4g22280"
CDS	complement(join(12761..13015,13081..13227,13379..14157, 14209..14314,14995..15111)) /note="similarity to hypothetical protein, chr.4, Arabidopsis thaliana, PATCHX:G3392763 contains EST gb:A1998861.1, A1993240.1" /codon_start=1 /product="putative protein" /protein_id="CAW9183.1" /db_xref="GI:7269074"
gene	/translacion="MEKWDMDTCYVGRNPATIVNVNGLNISHOQMISLKLKCVIIF FGVVADFLRLITPERFIRIMQMRINETSYSLSMWRYLFAFRPLACDDODE VGCGSFDLPDRVLVTGNFPPIRKISKMSIDTGAVTRMWVDLEHGVSYYDIDD ISEDIGVPLEIFPCKTLVEKLARGFYANPDVSLPSLKTLELSIWCNDCCCP LGRIISACPVAEEELIIIGCWOHIEPCFTASSSTLKLTITSSEHYEWGITDPSP AYLESDDIPVRHPFVNESLVEAKIALDLCDSDSNPTNLIGLENVEYLESSGYSG ILYDREAIPLYSKLFRLSIKSNSRHYITMMLPLLIESPKLETLYIKGPLSDREAE REVJSCPVKEITEYGKXELEEOMHFLLKKLPCELVAVRASAINDKXSRTMD LLMYRSSNCINKKFC"
Intron	complement(13016..13080) /gene="AT4g22280"
exon	/number=1 complement(13081..13227) /gene="AT4g22280"
Intron	/number=2 complement(13228..13378) /gene="AT4g22280"
exon	/number=2 complement(13379..14157) /gene="AT4g22280"

8.2%; Score 39.4; DB 8; length 199577;

[illegible]

us-09-677-374-3.rge

Page 18

Search completed: October 27, 2002, 17:46:41  
Job time : 1387.41 secs

Search completed: October 27, 2002, 17:46:41  
Job time : 1387.41 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:15:49 : Search time 967.668 Seconds  
(without alignments)  
6736.842 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483  
Sequence: 1 atgcgtgctgctgctgcaggg.....aggctgattagaccgcaaaa 483

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:\*

1: em.estbda:\*  
2: em.esthum:\*  
3: em.estlin:\*  
4: em.estmnu:\*  
5: em.estov:\*  
6: em.estpl:\*  
7: em.estro:\*  
8: em.hlc:\*  
9: gb.estl:\*  
10: gb.estc2:\*  
11: gb.hlc:\*  
12: gb.gss:\*  
13: em.gss.hum:\*  
14: em.gss.iny:\*  
15: em.gss.pln:\*  
16: em.gss.vtc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	4.1	516	12	AQ518147 HS_5090_A
2	20	4.1	823	10	BI186311 UML-P-FN-
3	19	3.9	190	12	AL003232 F.rubrip
4	19	3.9	348	10	TS1729 YB28d01.r1
5	19	3.9	459	9	AV763598 AV763598
6	19	3.9	527	10	BM381932 BM381932
7	19	3.9	528	9	AM658421 AM658421
8	19	3.9	536	9	AI637152 AI637152
9	19	3.9	630	10	BM35231 MEST147-f
10	19	3.9	648	12	BM381604 AG-ND-132
11	19	3.9	697	10	BM350333 MEST264-C
12	19	3.9	744	10	BI764273 603045965
13	19	3.9	2405	10	BM381932 MEST542-A
14	18	3.7	372	12	BM381932 MEST542-A
15	18	3.7	386	9	AM145423 ga20h03.y
16	18	3.7	387	12	AZ480429 IM0301N22
17	18	3.7	393	10	BF656950 OV2_19_B0

18	18	3.7	394	12	AQ408806	AQ408806 HS_5115_A
19	18	3.7	405	10	BE552264	BE552264 hy05c03.x
20	18	3.7	405	12	B76017	B76017 RPT111-13EL
21	18	3.7	406	9	AI953147	AI953147 wq30h02.x
22	18	3.7	411	9	AI635553	AI635553 ts95e09.x
23	18	3.7	441	10	R98050	R98050 yr30d06.s1
24	18	3.7	445	9	BB748581	BB748581 BR748581
25	18	3.7	487	10	BI479759	BI479759 WHE3451.G
26	18	3.7	488	9	AI167944	AI167944 o009b08.x
27	18	3.7	509	10	BI326875	BI326875 AK070H081
28	18	3.7	516	10	BE405086	BE405086 WHE1215_E
29	18	3.7	523	9	AI628427	AI628427 ly76d03.x
30	18	3.7	524	12	BM371300	BM371300 AG-ND-131
31	18	3.7	533	9	AA884310	AA884310 am15c05.s
32	18	3.7	560	12	A29133217	A29133217 RPT-24-1
33	18	3.7	564	12	AQ378208	AQ378208 RPT-23-3
34	18	3.7	567	10	BI683491	BI683491 464898 MA
35	18	3.7	568	12	BM378459	BM378459 AG-ND-165
36	18	3.7	569	12	BM393470	BM393470 AG-ND-145
37	18	3.7	573	12	BM382024	BM382024 AG-ND-140
38	18	3.7	586	10	BM3712909	BM3712909 pg11n.pk0
39	18	3.7	587	12	BM394407	BM394407 AG-ND-163
40	18	3.7	600	10	T67705	T67705 yc27c07.s1
41	18	3.7	601	9	AI742638	AI742638 wq44d05.x
42	18	3.7	602	12	BH072402	BH072402 RPT-24-2
43	18	3.7	603	9	AV596360	AV596360 AV596360
44	18	3.7	616	9	AM950089	AM950089 EST362279
45	18	3.7	635	12	BH372181	BH372181 AG-ND-167

#### ALIGNMENTS

RESULT 1  
AQ518147  
LOCUS  
DEFINITION  
HS\_5090.AL.F08.T7A RPT-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=666 COL=15 Row=K, DNA sequence.  
VERSION  
AQ518147.1 GI:4743329  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 516)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Library availability, please contact Pieter de Jong  
(pieterdejong.med.bu@u.washington.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.bu@u.washington.edu/ordering\_bac.htm)  
or from Research Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 666 row: K column: 15  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 516.  
Location/Qualifiers  
1. 516  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
 /clone="plate=666 Col=15 Row=K"  
 /clone\_lib="RPI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACE3.6 vector at EcoRI sites"  
 BASE COUNT 122 a 156 c 129 g 104 t 5 others  
 ORIGIN

Query Match  
 Best Local Similarity 4.1%; Score 20; DB 12; Length 516;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 GCCTGAGGCGAGCTCTCTG 30  
 Db 191 GCCTGAGGCGAGCTCTCTG 210

## RESULT 2

B1186311/c

LOCUS

DEFINITION UNT-P-FN-cf-b-07-0-UNT.s2 UNT-P-FN Sus scrofa cDNA clone  
 B1186311  
 VERSION B1186311  
 KEYWORDS GI:14660720  
 SOURCE EST.  
 ORGANISM Pig.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE 1 (bases 1 to 823)  
 Caetano, A.R., Johnson, R.K. and Pomp, D.  
 Generation and sequence characterization of a normalized cDNA  
 library from swine ovarian follicles  
 Unpublished (2001)  
 CONTACT: Pomp, D  
 Department of Animal Science  
 University of Nebraska, Lincoln  
 Lincoln, NE 68583-0908, USA  
 Tel: 402 472 6416  
 Fax: 402 472 6342  
 Email: dpomp@unl.edu  
 Oligo-dT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message.  
 Seq primer: M13 -29  
 POLYA-No.

## JOURNAL

COMMENT

Location/Qualifiers  
 1. 823  
 /organism="Sus scrofa"  
 /strain="University of Nebraska, Lincoln Swine Selection  
 Lines"  
 /db\_xref="taxon:9823"  
 /clone="UNT-P-FN-cf-b-07-0-UNT"  
 /clone\_lib="UNT-P-FN"  
 /dev\_stage="ADULT"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UNT-P-FN  
 library is a normalized library representing porcine  
 ovarian follicles, ranging between 2.0 to 10.0 mm in  
 diameter, collected during 7 days of the follicular phase  
 of the pig estrous cycle. This library was derived from  
 the library UNT-P-F2. The tag is a string of 5-6  
 nucleotides present between the Not I site and the  
 oligo-dT track. The library was constructed as described  
 by Bonaldi, Lennon and Soares, Genome Research 6: 791-806  
 1996.

## FEATURES

source

BASE COUNT  
 ORIGIN

151 a 272 c 231 g 167 t 2 others  
 TAG-SEQ=None found"  
 ORIGIN

Query Match  
 Best Local Similarity 4.1%; Score 20; DB 10; Length 823;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 8 GTTCCGCGAGGCGAGCTCT 27  
 Db 730 GTTCCGCGAGGCGAGCTCT 711

## RESULT 3

FR0011976

LOCUS

DEFINITION F.rubripes GSS sequence, clone 072L05aE5, genomic survey sequence.  
 ACCESSION AL003232.1 GI:2448802  
 VERSION AL003232.1  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Takifugu rubripes.  
 ORGANISM Takifugu rubripes.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

REFERENCE 1 (bases 1 to 190)  
 Edgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umranaia, Y.,  
 Williams, G. and Brenner, S.  
 Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource  
 Centre Hinxton, Cambridge, CB10 1SB. Email: biolhelp@hgmrc.ac.uk  
 V-type: plasmid  
 PRIMER: M13  
 DESC: One pass dye-terminator sequencing of cosmid cloned genomic  
 sequence.

COMMENT

JOURNAL

TITLE

DIRECT SUBMISSION

SUBMITTED (09-SEP-1997)

VECTOR: M13mp18

PRIMER: M13

## FEATURES

source

Location/Qualifiers  
 1. 190  
 /organism="Takifugu rubripes"  
 /db\_xref="taxon:31033"  
 /clone\_lib="Cosmid 072L05"  
 /clone="072L05aE5"  
 BASE COUNT 56 a 38 c 40 g 56 t  
 ORIGIN  
 Query Match  
 Best Local Similarity 3.9%; Score 19; DB 12; Length 190;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 16 CAGGCGAGCTCTCTGATCA 34  
 Db 119 CAGGCGAGCTCTCTGATCA 137

## RESULT 4

T51729

LOCUS

DEFINITION yb28d01.r1 StrataGene fetal spleen (#937205) Homo sapiens cDNA  
 T51729  
 VERSION T51729  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 348)  
 Hillier, L., Lennon, G., Becker, M., Bonaldi, M.F., Chapelli, B.,  
 Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins,  
 M., Hiltman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore  
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,  
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags

TITLE

```

/organism="zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MESF542-A12"
/clone_lib="ISUM6"
/tissue_type="mixed"
/lab_host="DH10B"
/note="Vector: pSLiP7 (4.43 kb); Site_1: EcoRI; Site_2:
NotI. Tissue samples were collected and partially pooled
prior to RNA extraction. First-strand cDNAs were prepared
from 21 individual pools of oligo-dT selected mRNAs by
priming with 21 different NotI oligo-dT tag primers
(3'-AAGCTGAGAGATTTCGGCGCCGNNNNNNTTTTTTTTTTTT-3').
Distinguishable 'bar code' tags, (N)6, were used for each.

```

separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are:

ATACGC--Germinated seeds and seedlings (1, 2, 8, 11 DAG);  
 ATACGC--Mixed mature tissues (17, 21, 38, 69, 77 DAG);  
 CACAGC--Kernels (3, 5, 10, 15, 20, 25, 30, DAP);  
 TACACC--Adventitious roots (65 DAG);  
 CACGCG--Tassels (3-39 cm, 53 and 56 DAG);  
 AGTAC--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG);  
 TGAGCC--Husks (73 DAG);  
 GACGAC--Silks;  
 AATCGG--unpollinated first ears; CTAAGC--ear shanks;  
 GTCGAC--etiolated seedlings; ACTGAG--callus;  
 GTCGAC--Cytoloximide-treated callus; GTCGAC--anaerobic treated seedlings; CTRCA--NAA (a-Naphthalene acetic acid) treated seedlings; GATGCG--Kinetin-treated seedlings; MAGACC--ACPC (1-aminocyclopropane-1-carboxylic acid) treated seedlings; GCTTCA--Brassinolide-treated seedlings; CTRAGC--ABA (Abscissic acid)-treated seedlings; TACGCA--GA (Gibberellic acid)-treated seedlings; GCAGCA--JA (Jasmonic acid)-treated seedlings; Equal amounts of first-strand cDNA from each reaction were combined and used as template for DNA polymerase II strand synthesis. After the addition of EcoRI adaptors, and 2.0 kb were directionally cloned into the EcoRI and NotI sites of the pSLIP7 expression vector. Plasmid DNA isolated from the library was digested with NotI to remove empty vector clones. Linear DNAs from 5.4 to 7 kb were gel purified and ligated at low concentration to promote recircularization. Ligation products were precipitated and transformed into DH10B host cells. The complexity of the resulting library was 1.2 x 10<sup>6</sup>.

BASE COUNT 94 a 149 c 179 g 105 t  
 ORIGIN

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 10; Length 528;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 167 TCAGCGGTGGGTTCTGGG 185  
 Db 365 TCAGCGGTGGGTTCTGGG 383  
 |||||||||||||||||||

RESULT 7  
 LOCUS AM658421 528 bp mRNA linear EST 25-APR-2001  
 DEFINITION 94440 MARC BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION AM658421  
 VERSION AM658421.1 GI:7424247  
 KEYWORDS EST.  
 SOURCE cow.  
 ORGANISM Bos taurus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.  
 Smith, T.P.L., Grosse, W.M., Preking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karimychewa, S., Hwang, F., Quackenbush, J. and Keefe, J.W.

TITLE  
 JOURNAL MEDLINE  
 COMMENT Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -m1nscore 18 and -mismatch 12 options.

PCR PRIMERS  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCAGCTACGAGC  
 Plate: 62 row: p column: 3  
 Seq primer: ATTATGTCACATATGAC.  
 Location/Qualifiers

FEATURES  
 source 1..528  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV SPOR6; Site:1: XbaI; Site:2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 97 a 164 c 150 g 117 t  
 ORIGIN

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 9; Length 528;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 211 CAGAGCATGAGCAGCAGCAGG 229  
 Db 67 CAGAGCATGAGCAGCAGCAGG 49  
 |||||||||||||||||||

RESULT 8  
 LOCUS A1637152 536 bp mRNA linear EST 26-APR-1999  
 DEFINITION 603001D08.x1 603 - stressed root cDNA library from Wang/Bohnert lab  
 ACCESSION A1637152  
 VERSION A1637152  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.

REFERENCE  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade: Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 536)  
 Walbot, V.

TITLE  
 JOURNAL  
 COMMENT Maize ESTs from various cDNA libraries sequenced at Stanford  
 Unpublished (1999)  
 Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu

FEATURES  
 source 1..536  
 Location/Qualifiers

1..536  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="603 - stressed root cDNA library from Wang/Bohnert lab"  
 /tissue\_type="seedling"  
 /dev\_stage="salt stress"  
 /lab\_host="E. coli XL Gold"  
 /note="Organ: root; Vector: pBluescriptII SK(+); XR; Seedling stressed root cDNA library from Wang/Bohnert lab"

BASE COUNT 95 a 150 c 195 g 94 t  
 ORIGIN

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 9; Length 536;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 167 TCGGCGTGGCGTTCTGGG 185  
 |||||||  
 Db 329 TCGGCGTGGCGTTCTGGG 347

RESULT 9  
 BM335231  
 LOCUS  
 DEFINITION BM335231 630 bp mRNA linear EST 16-JAN-2002  
 MEST147-F07.T3 ISUM5-RN Zea mays cDNA clone MEST147-F07 3, mRNA  
 sequence.  
 ACCESSION BM335231  
 VERSION BM335231.1 GI:18165392  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 630)  
 Men.T.J., Qiu.F., Guo.L., Ashlock, D.A and Schnable, P.S.  
 REFERENCED: Sequence Tags from B73 Maize: Various stages and tissues  
 including seedlings treated with a variety of hormones  
 Unpublished (2001)  
 JOURNAL Contact: Patrick S. Schnable  
 COMMENT Schnable Laboratory  
 Iowa State University  
 6405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA  
 Tel: 515-294-0975  
 Fax: 515-294-2299  
 Email: schnable@iastate.edu  
 Individual basecall and confidence value were assigned using the  
 phred software.  
 (<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b  
 rt). Overall sequence quality assessment and vector trimming were  
 conducted using the Lucy software (<http://www.tigr.org/soflab/>).  
 Lucy parameters were set to ensure an overall trimmed quality of  
 97.5% or better without any vector fragments in the chosen  
 high-quality region of each sequence. Low-quality bases between the  
 poly-T and the high-quality region were replaced with N's to serve  
 as spacers.  
 PCR Primers  
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
 Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).  
 FEATURES  
 source  
 Location/Qualifiers  
 1..630  
 /organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="MEST147-F07"  
 /clone\_1lb="ISUM5-RN"  
 /tissue\_type="mixed"  
 /lab\_host="DH108"  
 /note="Vector: pTR3PAC; Site\_1: EcoRI; Site\_2: NotI;  
 tissues: Germinated seed and seedlings (1', 2', 8', 11 DAG),  
 mixed mature tissues (17', 21', 38', 69', 77 DAG), kernels  
 (3', 5', 10', 15', 20', 25', 30', 36P), Adventitious roots (65 DAG  
 ), tassels (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0  
 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated  
 first ear, ear shank, etiolated seedlings, callus,  
 cycloheximide-treated callus, Anaerobic treated seedlings  
 cycloheximide-treated acetec acid)-treated seedlings,  
 Kinetin-treated seedlings, ACP  
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
 brassinolide-treated seedlings, ABA (Abscissic acid  
 )-treated seedlings, GA (Gibberellic acid)-treated  
 seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA  
 molecules were generated as follows: First-strand cDNA was  
 prepared from oligo-dT selected mRNA by priming with a  
 NotI oligo-dT primer (5'  
 AACTGGAGAATTCGGCGCCGAGGAATTTTATTTTATTTTATTTT). The  
 resulting DNA:RNA hybrid was treated with RNase H and used

as a template for DNA Pol-I-catalyzed second strand  
 synthesis. After the addition of EcoRI adaptors, the  
 ds-cDNAs were digested with NotI and size-selected. The  
 resulting molecules were directionally cloned into the  
 EcoRI and NotI sites of the pTR3PAC vector. The library  
 then went through one round of normalization to Cot value  
 of 5 based on the methods of Marcello Bento Soares (Genome  
 Research 6: 791-806, 1996)."

BASE COUNT 102 a 187 c 228 g 113 t  
 ORIGIN

Query Match 3.9%; Score 19; DB 10; Length 630;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 167 TCGGCGTGGCGTTCTGGG 185  
 |||||||  
 Db 352 TCGGCGTGGCGTTCTGGG 370

RESULT 10  
 BH381604  
 LOCUS  
 DEFINITION BH381604 648 bp DNA linear GSS 10-DEC-2001  
 AG-ND-132N21.TR ND-TAM Anopheles gambiae genomic clone AG-ND-132N21  
 , DNA sequence.  
 ACCESSION BH381604  
 VERSION BH381604.1 GI:17327746  
 KEYWORDS GSS.  
 SOURCE African malaria mosquito.  
 ORGANISM Anopheles gambiae  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae  
 ; Anophelinae.  
 1 (bases 1 to 648)  
 Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.  
 REFERENCED: Direct Submission of BAC-end sequences from Anopheles gambiae  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: AG-ND-132N21.TR  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjoftus@tigr.org  
 This clone is from an A. gambiae BAC library (ND-TAM) provided by  
 F.H. Collins and sequenced by The Institute for Genomic Research  
 (TIGR). The BAC library was generated from A. gambiae PEST strain  
 DNA. All DNA was extracted from newly hatched first instar larvae  
 to minimize the inclusion of DNA from microorganisms that inhabit  
 the gut. The DNA is derived from mixed sexes of larvae. The BAC  
 library was constructed at Texas A&M University BAC Center  
 University, College Station, Texas 77843-2123, USA using a HindIII  
 partial digest.  
 Seq primer: M13 For  
 Class: BAC ends.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..648  
 /organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-132N21"  
 /clone\_1lb="ND-TAM"  
 /note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 168 a 132 c 114 g 234 t  
 ORIGIN

Query Match 3.9%; Score 19; DB 12; Length 648;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 252 TCTGGAAGAAGTGAAGCC 270  
 |||||||

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 697)  
Men, T. J., Qiu, F., Guo, L., Ashlock, D. A and Schnable, P. S.  
Expressed Sequence Tags from B73 Maize: various stages and tissues  
including seedlings treated with a variety of hormones  
Unpublished (2001)  
Contact: Patrick S. Schnable

PCR Primers:  
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
 Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).  
 Location/Qualifiers  
 1 607

/organism="Zea mays"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone\_name="MES764-C11"  
 /clone\_lib="ISUM5-RN"  
 /tissue\_type="mixed"  
 /lab\_host="DH10B"  
 /note="Vector: pTRT3PAC; Site.1: EcoRI; Site.2: NotI.  
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), kernels  
 (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG),  
 tassels (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.3-  
 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated  
 first ear, ear shank, etiolated seedlings, callus,  
 Cycloheximide-treated callus, Anaerobic treated seedlings,  
 MA (α-Naphthalene acetic acid)-treated seedlings,  
 Kinetin-treated seedlings, ACPG  
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
 Brassinolide-treated seedlings, ABA (Abscissic acid  
 )-treated seedlings, GA (Gibberellic acid)-treated  
 seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
 molecules were generated as follows. First-strand cDNA was  
 prepared from oligo-dT selected mRNA by priming with a  
 NotI oligo-dT primer (5'  
 AATCGAAGATTCCGCCGCCGACGAAATTTTTTTTTTTTTT). The  
 resulting DNA:RNA hybrid was treated with RNase H and used  
 as a template for DNA PolI-catalyzed second strand  
 synthesis. After the addition of EcoRI adaptors, the  
 ds-cDNAs were digested with NotI and size-selected. The

Query Match	3.98;	Score 19;	DB 10;	Length 697;
Best Local Similarity	100.08;	Pred. No. 42;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Score 2

<b>DY</b>	167	TGCGCGGTGGGTTCTGGG	185
<b>Db</b>	370	TGCGCGGTGGGTTCTGGG	388

RESULT	12
BI764273	
LOCUS	744 bp mRNA linear EST 25-SEP-2001
DEFINITION	6030459565f1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5186200 5',
ACCESSION	BI764273
VERSION	BI764273.1 GI:15755851
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE  
Mammalia; Eutheria: Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 744)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Sprussberry, ph.d.  
COMMENT

CDNA library preparation: Life technologies, Inc.  
CDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
plate: LLNL1465 row: a column: 17  
High quality sequence stop: 692.  
location/quality:

```

FEATURES
source
Location/Qualifiers:
1..744
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5186200"
/clone_1ib="NIH_MGC_116"
/lab_host="DH10B"
Note="Organ: pooled colon, kidney, stomach. Vector:
pCMV-SPORT6. Site_1: NotR. site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon, 46 yo male kidney, and pool of
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
BASE COUNT
147 a 238 c 238 g 121 t

```

Query Match	3.9%	Score 19;	DB 10;	Length 744;
Best Local Similarity	100.0%	Pred. No. 43;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 259	AAAGTGAAGCGGCACG	277		
Db 348	AAAGTGAAGCGGCACG	366		

RESULT 13  
BG715056/c  
LOCUS  
DEFINITION BG715056 2405 bp mRNA linear EST 08-MAY-2001  
602673281F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4795701 5',  
mRNA sequence.  
ACCESSION BG715056  
VERSION BG715056.1 GI:13993987  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2405)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LHAM0679 row: b column: 22  
High quality sequence stop: 5.  
Location/Qualifiers  
1..2405  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="4795701"  
/clone\_lib="NIH\_MGC\_96"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-xhoI (ctcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.3 kb and  
normalized to 10^5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 554 a 1303 c 201 g 347 t  
ORIGIN

Query Match 3.9%; Score 19; DB 10; Length 2405;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GTGGCGTTGTGGCGGTCT 119  
|||||  
Db 2213 CTGGCGTTGTGGCGGTCT 2195

RESULT 14  
BH395820 372 bp DNA linear GSS 11-DEC-2001  
LOCUS  
DEFINITION AG-ND-119B5.TF ND-TAM Anopheles gambiae genomic clone AG-ND-119B5,  
DNA sequence.  
ACCESSION BH395820  
VERSION BH395820.1 GI:17341961  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae  
; Anopheles.  
1 (bases 1 to 372)  
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE  
JOURNAL  
COMMENT Direct Submission of BAC-end sequences from Anopheles gambiae  
Unpublished (2001)  
Other\_GSSs: AG-ND-119B5.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: [bjloftus@ligr.org](mailto:bjloftus@ligr.org)  
This clone is from an A. gambiae BAC library (ND-TAM) provided by  
F.H. Collins and sequenced by The Institute for Genomic Research  
(TIGR). The BAC library was generated from A. gambiae PEST strain  
DNA. All DNA was extracted from newly hatched first instar larvae  
to minimize the inclusion of DNA from microorganisms that inhabit  
the gut. The DNA is derived from mixed sexes of larvae. The BAC  
library was constructed at Texas A&M University BAC Center  
University, College Station, Texas 77843-2123, USA using a HindIII  
partial digest.  
Seq primer: M13 For  
Class: BAC ends.  
Location/Qualifiers  
1..372  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone\_image="AG-ND-119B5"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 90 a 84 c 66 g 132 t  
ORIGIN

Query Match 3.7%; Score 18; DB 12; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1,3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 CTGCAAAAGTGAAGGCC 270  
|||||  
Db 142 CTGCAAAAGTGAAGGCC 159

RESULT 15  
AW145423 386 bp mRNA linear EST 01-NOV-1999  
LOCUS  
DEFINITION g620H03.Y1 Moss EST library PPU Physcomitrella patens cDNA clone  
PEP\_SOURCE\_ID:PPU021606 5', mRNA sequence.  
ACCESSION AW145423  
VERSION AW145423.1 GI:6167159  
KEYWORDS EST.  
SOURCE Physcomitrella patens.  
ORGANISM Physcomitrella patens.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
1 (bases 1 to 386)  
Quatrano,R., Bashardes,S., Cove,D., Cumling,A., Knight,C., Clifton  
,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wyllie,T., Underwood  
,K., Theising,B., Allen,K., Bowers,Y., Person,B., Swaller,T.,  
Stephens,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,  
Waterston,R. and Wilson,R.  
Leeds/Mash U Moss EST Project  
Unpublished (1999)  
Contact: Ralph Quatrano  
Leeds/Mash U Moss EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
Libraries were constructed by Dr. Stavros Bashardes as part of the  
Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and  
Washington Univ. in St. Louis (USA) DNA sequencing by: Washington  
University Genome Sequencing Center For information on obtaining a  
clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco  
 High quality sequence stop: 373.  
 Location/Qualifiers  
 1. 386

/organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 /clone="PEP\_SOURCE\_ID:PPU021606"  
 /clone\_lib="Moss EST library ppy"  
 /tissue\_type="protonemata: 7 day old tissue  
 ammonium-grown"  
 /lab\_host="DH10B"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Construction of the cDNA library was carried out  
 using Stratagene's 'UnizAP' - cDNA synthesis kit'. cDNA  
 was constructed using an oligo dt primer/linker that  
 contains a XhoI site within it. Following ds cDNA  
 synthesis, EcoRI adapters were ligated to the blunt ends  
 and sample was digested with XhoI. The result is cDNA  
 with an EcoRI sticky end on one side and a XhoI sticky  
 end on the other. This cDNA was ligated directionally in  
 pUnizAP arms. The vector is designed containing the  
 pBluescript sequence as well as lambda DNA and cDNA is  
 then packaged using Gold gigapackaging extracts. Library  
 was grown in XLIRBlue MRF' cells and amplified. The library  
 was excised by mass excision using Stratagene's 'Mass  
 excision kit' that uses exassist as a helper phage that  
 releases the pBluescript sequence and circularises it as  
 single stranded plasmids that are then packaged (by helper  
 phage) and secreted out of the host cell as phagemids.  
 SOUR cells were transformed with phagemids and the library  
 was plated out on LB-amp plates to select for  
 transformants. Approximately 1,000,000 colonies were grown  
 and recovered. The double stranded plasmid library was  
 recovered by using Qulagen Midi prep kit. 2 micro grams of  
 each library were used to transform DH10B cells by  
 electroporation."

BASE COUNT 104 a 82 c 113 g 87 t  
 ORIGIN

## Query Match

Best Local Similarity 3.7%; Score 18; DB 9; Length 386;  
 Matches 18; Conservativity 100.0%; Pred.No. 1.3e+02;

Mismatches 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 GTTGCCCTGCAGGCGACCT 25  
 ||||||||||||||||  
 Db 187 GTTGCCCTGCAGGCGACCT 170

Search completed: October 27, 2002, 22:36:41  
 Job time : 972.668 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:42:44 : Search time 248.035 Seconds  
(without alignments)  
7028.450 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483  
Sequence: 1 atgcgtgctgctgctgcaagg.....aggtgattagcaccgaaaaa 483

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2891707 seqs, 1804656691 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5783414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Pending\_Patents\_NA\_New:\*  
1: /cgn2\_6/prodata/1/pna/PCT\_NEW\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/pna/US06\_NEW\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/pna/US10\_NEW\_COMB.seq:\*  
7: /cgn2\_6/prodata/1/pna/US10\_NEW\_COMB.seq2:\*  
8: /cgn2\_6/prodata/1/pna/US60\_NEW\_COMB.seq3:\*  
9: /cgn2\_6/prodata/1/pna/US60\_NEW\_COMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	4.6	12848	US-10-114-170-252	Sequence 252, App
2	22	4.6	38155	US-10-114-170-79	Sequence 79, Appl
3	22	4.6	39824	US-10-053-853A-876	Sequence 876, App
4	22	4.6	48423	US-10-053-853A-484	Sequence 484, App
5	22	4.6	48908	US-10-114-170-137	Sequence 137, App
6	22	4.6	58175	US-10-053-853A-137	Sequence 337, App
7	22	4.6	91772	US-10-053-853A-1	Sequence 1, Appl
8	19	3.9	453	US-09-918-995-216	Sequence 216, App
9	19	3.9	455	US-09-919-002-8733	Sequence 8373, App
10	19	3.9	847	US-10-211-818-27	Sequence 27, Appl
11	19	3.9	2270	PCT-US02-2632-47	Sequence 47, Appl
12	19	3.9	2332	US-10-219-999-30827	Sequence 30827, A
13	19	3.9	2329	US-10-108-260A-2343	Sequence 2343, Ap
14	19	3.9	2329	US-10-108-260-2343	Sequence 2343, Ap
15	18	3.7	440	US-10-221-279-4847	Sequence 4847, Ap
16	18	3.7	466	US-09-918-995-8784	Sequence 8784, Ap
17	18	3.7	643	US-10-027-632-276565	Sequence 276565,
18	18	3.7	643	US-10-027-632-276567	Sequence 276567,
19	18	3.7	644	US-10-027-632-276566	Sequence 276566,
20	18	3.7	705	US-10-027-632-173497	Sequence 173497,
21	18	3.7	705	US-10-027-632-173498	Sequence 173498,
22	18	3.7	716	US-10-027-632-22462	Sequence 22462, A
23	18	3.7	1300	PCT-US02-25766-2948	Sequence 2948, Ap
24	18	3.7	1855	US-10-027-632-261467	Sequence 261467,

C 25	18	3.7	1855	US-10-027-632-261468	Sequence 261468,
C 26	18	3.7	2661	US-60-360-039-32215	Sequence 32215, A
C 27	18	3.7	2690	US-60-360-039-28064	Sequence 28064, A
C 28	18	3.7	2690	US-60-360-039-30818	Sequence 30818, A
C 29	18	3.7	11736	US-09-663-870A-1	Sequence 1, Appl
30	17	3.5	123	US-10-182-995-11452	Sequence 11452, A
31	17	3.5	123	US-10-203-134-14602	Sequence 14602, A
32	17	3.5	123	US-10-203-136-14617	Sequence 14617, A
33	17	3.5	123	US-10-182-993-14306	Sequence 14306, A
34	17	3.5	123	US-10-182-997-10762	Sequence 10762, A
35	17	3.5	123	US-10-182-998-6531	Sequence 6531, Ap
36	17	3.5	123	US-10-203-135-14118	Sequence 14118, A
37	17	3.5	123	US-10-203-137-14750	Sequence 14750, A
38	17	3.5	123	US-10-203-138-6710	Sequence 6710, Ap
39	17	3.5	123	US-10-203-139-14205	Sequence 14205, A
40	17	3.5	280	US-09-940-210B-2218	Sequence 2218, Ap
C 41	17	3.5	359	US-09-721-544-23832	Sequence 23832, A
C 42	17	3.5	390	US-10-155-881-6701	Sequence 6701, Ap
C 43	17	3.5	451	US-09-918-995-22279	Sequence 22279, A
44	17	3.5	459	US-09-918-995-26505	Sequence 26505, A
45	17	3.5	486	US-10-182-995-1547	Sequence 1547, Ap

#### ALIGNMENTS

RESULT 1  
US-10-114-170-252/c  
Sequence 252, Application US/10114170  
GENERAL INFORMATION:  
APPLICANT: Blattner, Frederick R.  
Perna, Valerie  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: Novel Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
/CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/114,170  
FILING DATE: 01-Apr-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/453,702  
FILING DATE: 03-DEC-1999  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296,95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ. ID NO: 252:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12848  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 252:

US-10-114-170-252

Query Match 4.6%; Score 22; DB 8; Length 12848;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAAATCCGGACACCGGTAAACAGC 312  
Db 3851 TAAATCCGGACACCGGTAAACAGC 3830

## RESULT 2

US-10-114-170-79  
Sequence 79, Application US/10114170  
GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.  
Burland, Valerie  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: Novel Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.44mb storage

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/114,170  
FILING DATE: 01-Apr-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/453,702  
FILING DATE: 03-DEC-1999  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38155  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-10-114-170-79  
SEQUENCE DESCRIPTION: SEQ ID NO: 79:

Query Match 4.6%; Score 22; DB 8; Length 38155;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAAATCCGGACACCGGTAAACAGC 312  
Db 31926 TAAATCCGGACACCGGTAAACAGC 31947

## RESULT 3

US-10-053-853A-876  
Sequence 876, Application US/10053853A  
GENERAL INFORMATION:

```

; TITLE OF INVENTION: Novel Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48908
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-10-114-170-137

Query Match          4.6%; Score 22; DB 8; Length 48908;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAACAGC 312
Db 15032 TAATCCGACACCGGTAACAGC 15011

RESULT 6
US-10-053-853A-337
; Sequence 337, Application US/10053853A
; GENERAL INFORMATION:
; APPLICANT: HAYASHI, Hideo
; APPLICANT: SHINAGAWA, Hideo
; APPLICANT: MAKINO, Kozo
; APPLICANT: HAYASHI, Tetsuya
; APPLICANT: OHNISHI, Makoto
; APPLICANT: HATTORI, Masahira
; APPLICANT: KUROKAWA, Ken
; TITLE OF INVENTION: Polynucleotide molecules and polypeptides specific to Enterohemorrhagic E. coli O157:H7 and use thereof
; FILE REFERENCE: 2002-0060A/MMC/01704
; CURRENT APPLICATION NUMBER: US/10/053,853A
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: JP2001-112010
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 1866
; SEQ ID NO 337
; LENGTH: 58175
; TYPE: DNA
; ORGANISM: Escherichia coli O157:H7
US-10-053-853A-337
```

```

Query Match          4.6%; Score 22; DB 8; Length 58175;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAACAGC 312
Db 31033 TAATCCGACACCGGTAACAGC 31054

RESULT 7
US-10-053-853A-1/c
; Sequence 1, Application US/10053853A
; GENERAL INFORMATION:
; APPLICANT: HAYASHI, Hideo
; APPLICANT: SHINAGAWA, Hideo
; APPLICANT: MAKINO, Kozo
; APPLICANT: HAYASHI, Tetsuya
; APPLICANT: OHNISHI, Makoto
; APPLICANT: HATTORI, Masahira
; APPLICANT: KUROKAWA, Ken
; TITLE OF INVENTION: Polynucleotide molecules and polypeptides specific to Enterohemorrhagic E. coli O157:H7 and use thereof
; FILE REFERENCE: 2002-0060A/MMC/01704
; CURRENT APPLICATION NUMBER: US/10/053,853A
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: JP2001-112010
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 1866
; SEQ ID NO 1
; LENGTH: 91772
; TYPE: DNA
; ORGANISM: Escherichia coli O157:H7
US-10-053-853A-1

Query Match          4.6%; Score 22; DB 8; Length 91772;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAACAGC 312
Db 66951 TAATCCGACACCGGTAACAGC 66930

RESULT 8
US-09-918-995-216
; Sequence 216, Application US/09918995
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 216
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(453)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-216

Query Match          3.9%; Score 19; DB 5; Length 453;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AAAGTGAAGCGGCGCAGC 277
```

Db

100 AAAGTGAAGCCGCCAGG 118

## RESULT 9

US-09-919-002-8373  
Sequence 8373, Application US/09919002  
GENERAL INFORMATION:  
APPLICANT: Leeshowitz, Dena  
APPLICANT: Liu, Jin  
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA  
FILE REFERENCE: 20411-752CON1  
CURRENT APPLICATION NUMBER: US/09/919,002  
PRIOR FILING DATE: 2001-07-30  
PRIOR FILING DATE: FILING DATE: 1999-07-22  
PRIOR FILING DATE: FILING DATE: 1998-02-13  
NUMBER OF SEQ ID NOS: 13203  
SOFTWARE: fastseq for Windows Version 3.0  
SEQ ID NO 8373  
LENGTH: 455  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(455)  
OTHER INFORMATION: n = A,T,C or G  
US-09-919-002-8373

## Query Match

Best Local Similarity 3.9%; Score 19; DB 5; Length 455;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AAAGTGAAGCCGCCAGG 277  
Db 378 AAAGTGAAGCCGCCAGG 396

## RESULT 10

US-10-211-818-27  
Sequence 27, Application US/10211818  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM030C1N  
CURRENT APPLICATION NUMBER: US/10/211,818  
PRIOR FILING DATE: 2002-08-05  
PRIOR FILING DATE: 2001-01-11  
PRIOR FILING DATE: 2002-08-05  
PRIOR FILING DATE: 2000-01-31  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 722  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 27  
LENGTH: 847  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (717)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (793)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (804)  
OTHER INFORMATION: n equals a,t,g, or c

## FEATURE:

NAME/KEY: misc\_feature  
LOCATION: (812)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (833)  
OTHER INFORMATION: n equals a,t,g, or c  
US-10-211-818-27

## Query Match

Best Local Similarity 3.9%; Score 19; DB 8; Length 847;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AAAGTGAAGCCGCCAGG 277  
Db 290 AAAGTGAAGCCGCCAGG 308

## RESULT 11

PCT-US02-26322-47

Sequence 47, Application PC/TUS0226322  
GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.

APPLICANT: YUE, Henry

APPLICANT: LU, Dying Anna M.

APPLICANT: SWARNAKAR, Anita

APPLICANT: TRANG, Y. Tom

APPLICANT: GRIFFIN, Jennifer A.

APPLICANT: EMERLING, Brooke M.

APPLICANT: FORSTYHE, Ian J.

APPLICANT: YAO, Monique G.

APPLICANT: RAMKUMAR, Jayalaxmi

APPLICANT: RICHARDSON, Thomas W.

APPLICANT: BECHA, Shengya D.

APPLICANT: LEE, Ernestine A.

APPLICANT: WARREN, Bridget A.

APPLICANT: LEHR-MASON, Patricia M.

APPLICANT: BAUGHN, Mariah R.

APPLICANT: LI, Joana X.

APPLICANT: DUGGAN, Brendan M.

APPLICANT: GIETZEN, Kimberly J.

APPLICANT: LAL, Preeti G.

APPLICANT: BOROMSKY, Mark L.

APPLICANT: ISON, Craig H.

APPLICANT: THANGAVELU, Kavitha

APPLICANT: XU, Yuming

APPLICANT: LEE, Sally

APPLICANT: ELLIOTT, Vicki S.

APPLICANT: SPRAGUE, William W.

APPLICANT: AZIMZAI, Yalda

APPLICANT: HAFALIA, April J.A.

APPLICANT: DING, Li

APPLICANT: NGUYEN, Dannel B.

APPLICANT: HONCHELL, Cynthia D.

APPLICANT: LUO, Wen

APPLICANT: WALIA, Nandinder K.

APPLICANT: MAROUTS, Joseph

APPLICANT: JACKSON, Jennifer L.

APPLICANT: TRAN,uyen K.

TITLE OF INVENTION: INTRACELLULAR SIGNALING MOLECULES

FILE REFERENCE: PF-1145 PCP

CURRENT APPLICATION NUMBER: PCT/US02/26322

PRIOR FILING DATE: 2002-08-16

PRIOR FILING DATE: 2001-08-17

PRIOR FILING DATE: 2001-08-24

PRIOR FILING DATE: 2001-08-31

PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: US 60/326,390  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: US 60/328,952  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/345,468  
PRIOR FILING DATE: 2001-10-19  
PRIOR APPLICATION NUMBER: US 60/372,499  
PRIOR FILING DATE: 2002-04-12  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PERL Program  
SEQ ID NO 47  
LENGTH: 2270  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No: 3744219CB1  
PCT-US02-26322-47

Query Match 3.9%: Score 19; DB 1; Length 2270;  
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AAGTGAAGCCGCCAGC 277  
|||||  
Db 912 AAGTGAAGCCGCCAGC 930

RESULT 12  
US-10-219-999-30827

Sequence 30827, Application US/10219999

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Edgerton, Michael D

APPLICANT: Hinkle, Gregory J.

APPLICANT: Kovalic, David K.

APPLICANT: Liu, Jindong

APPLICANT: Stein, Joshua

TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT

FILE REFERENCE: 38-10(52726)C

CURRENT APPLICATION NUMBER: US/10/219,999

CURRENT FILING DATE: 2002-08-15

PRIOR APPLICATION NUMBER: US 60/324,109

PRIOR FILING DATE: 2001-09-21

PRIOR APPLICATION NUMBER: US 60/312,544

PRIOR FILING DATE: 2001-08-15

NUMBER OF SEQ ID NOS: 63520

SEQ ID NO 30827

LENGTH: 2322

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

NAME/KEY: CDS

LOCATION: (1571)..(2227)

OTHER INFORMATION:

US-10-219-999-30827

Query Match 3.9%: Score 19; DB 8; Length 2322;  
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 CAGAACTTCAGCCGCCAGC 82  
|||||  
Db 104 CAGAACTTCAGCCGCCAGC 122

RESULT 13  
US-10-108-260A-2343

Sequence 2343, Application US/10108260A

GENERAL INFORMATION:

APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: Novel full length cDNA

FILE REFERENCE: H1-A0106

CURRENT APPLICATION NUMBER: US/10/108,260A

CURRENT FILING DATE: 2002-03-27

NUMBER OF SEQ ID NOS: 3458

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2343

LENGTH: 2329

TYPE: DNA

ORGANISM: Homo sapiens

US-10-108-260A-2343

Query Match 3.9%: Score 19; DB 6; Length 2329;  
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AAGTGAAGCCGCCAGC 277  
|||||  
Db 988 AAGTGAAGCCGCCAGC 1006

RESULT 14  
US-10-108-260-2343

Sequence 2343, Application US/10108260

GENERAL INFORMATION:

APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: Novel full length cDNA

FILE REFERENCE: H1-A0106

CURRENT APPLICATION NUMBER: US/10/108,260

CURRENT FILING DATE: 2002-03-27

NUMBER OF SEQ ID NOS: 5458

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2343

LENGTH: 2329

TYPE: DNA

ORGANISM: Homo sapiens

US-10-108-260-2343

Query Match 3.9%: Score 19; DB 8; Length 2329;  
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 AAGTGAAGCCGCCAGC 277  
|||||  
Db 988 AAGTGAAGCCGCCAGC 1006

RESULT 15  
US-10-221-279-4847/C

Sequence 4847, Application US/10221279

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc

TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides

FILE REFERENCE: 21272-046

CURRENT APPLICATION NUMBER: US/10/221,279

CURRENT FILING DATE: 2002-09-06

PRIOR APPLICATION NUMBER: 09/574,454

PRIOR FILING DATE: 2000-05-19

PRIOR APPLICATION NUMBER: 09/519,705

PRIOR FILING DATE: 2000-03-07

NUMBER OF SEQ ID NOS: 12360

SOFTWARE: Custom

SEQ ID NO 4847

LENGTH: 440

TYPE: DNA

ORGANISM: Homo sapiens

US-10-221-279-4847

Query Match 3.7%: Score 18; DB 8; Length 440;  
Best Local Similarity 100.0%; Pred. No. 18;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 TGCAAGGCAGCTCTCTGA 31  
|||||

Tue Oct 29 08:31:00 2002

us-09-677-374-3.oligo.inpn

Page 6

Db 128 TGCAGGGCAGCTCTCTGA 111

Search completed: October 28, 2002, 00:39:40  
Job time : 375.035 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:19:49 : Search time 35.8705 Seconds  
(without alignments)  
3307.484 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483

Sequence: 1 atcgctggtgctgctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: OLIGO\_NUC

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA :  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq : \*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq : \*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq : \*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq : \*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq : \*  
6: /cgn2\_6/prodata/2/ina/backfileseq1.seq : \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	3.7	3808	2 US-08-916-917-3	Sequence 3, Appl
2	18	3.7	3808	2 US-08-972-631-3	Sequence 3, Appl
3	18	3.7	3808	2 US-08-972-629-3	Sequence 3, Appl
4	18	3.7	3808	2 US-08-972-630-3	Sequence 3, Appl
5	18	3.7	3808	2 US-08-672-211-3	Sequence 3, Appl
6	18	3.7	3808	3 US-09-225-170-3	Sequence 3, Appl
7	18	3.7	4411529	4 US-09-103-840A-1	Sequence 1, Appl
8	17	3.5	754	4 US-09-541-941B-26	Sequence 26, Appl
9	17	3.5	1384	1 US-08-553-888A-1	Sequence 1, Appl
10	17	3.5	1466	1 US-08-553-888A-2	Sequence 2, Appl
11	16	3.3	1566	4 US-09-032-742-1	Sequence 1, Appl
12	16	3.3	1566	4 US-09-032-742-12	Sequence 12, Appl
13	16	3.3	1566	4 US-09-032-742-13	Sequence 13, Appl
14	16	3.3	1566	4 US-09-032-742-15	Sequence 15, Appl
15	16	3.3	1566	4 US-09-032-742-16	Sequence 16, Appl
16	16	3.3	1566	4 US-09-032-742-18	Sequence 18, Appl
17	16	3.3	1566	4 US-09-032-742-19	Sequence 19, Appl
18	16	3.3	1879	2 US-08-403-852D-5	Sequence 5, Appl
19	16	3.3	1879	4 US-08-510-646B-5	Sequence 5, Appl
20	16	3.3	1879	4 US-09-231-818-5	Sequence 5, Appl
21	16	3.3	1896	3 US-08-501-572-6	Sequence 6, Appl
22	16	3.3	1896	3 US-09-040-444-6	Sequence 6, Appl
23	16	3.3	1921	1 US-08-164-839-28	Sequence 28, Appl
24	16	3.3	1921	1 US-08-583-799-28	Sequence 28, Appl
25	16	3.3	1921	1 US-09-232-200-68	Sequence 68, Appl
26	16	3.3	1998	4 US-09-232-197-68	Sequence 68, Appl
27	16	3.3	1998	4 US-09-232-201-68	Sequence 68, Appl

c	28	16	3.3	2019	4 US-09-063-950-3	Sequence 3, Appl
	29	16	3.3	2087	4 US-09-232-191-6	Sequence 6, Appl
	30	16	3.3	2087	4 US-09-232-200-6	Sequence 6, Appl
	31	16	3.3	2087	4 US-09-232-197-6	Sequence 6, Appl
	32	16	3.3	2087	4 US-09-232-201-6	Sequence 6, Appl
	33	16	3.3	2852	4 US-09-063-950-1	Sequence 1, Appl
	34	16	3.3	3173	3 US-08-468-856A-1	Sequence 1, Appl
	35	16	3.3	3173	3 US-08-468-856A-1	Sequence 1, Appl
	36	16	3.3	3407	1 US-08-253-155A-7	Sequence 7, Appl
	37	16	3.3	3461	3 US-08-468-856B-6	Sequence 6, Appl
	38	16	3.3	3461	3 US-08-468-856B-6	Sequence 6, Appl
	39	16	3.3	3962	2 US-08-658-665-189	Sequence 189, App
	40	16	3.3	4092	2 US-08-469-537A-106	Sequence 106, App
	41	16	3.3	4695	2 US-08-231-193A-57	Sequence 57, Appl
	42	16	3.3	4695	2 US-08-486-273A-57	Sequence 57, Appl
	43	16	3.3	4695	3 US-08-940-086A-57	Sequence 57, Appl
	44	16	3.3	4695	4 US-08-940-035A-57	Sequence 57, Appl
	45	16	3.3	5302	2 US-08-658-665-190	Sequence 190, App

#### ALIGNMENTS

RESULT 1  
US-08-916-917-3

Sequence 3, Application US/08916917  
Patent No. 5856132

GENERAL INFORMATION:

APPLICANT: Stephens, Len

APPLICANT: Braselmann, Sylvia

TITLE OF INVENTION: G-BETA-GAMMA REGULATED  
PHOSPHATIDYLINOSITOL-3' KINASE

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas  
City: New York

STATE: NY  
COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM compatible  
SOFTWARE: FASTSEQ for Windows version 2.0b

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/916,917

FILING DATE: 15-AUG-1997  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/672,211

FILING DATE: 27-JUN-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.  
REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 8549-0006-999  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935  
TELEFAX: 650-493-5556

TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3808 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
US-08-916-917-3

Query Match 3.7%; Score 18; DB 2; Length 3808;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 TCAGAAACAGGAATCTA 428  
DB 2951 TCAGAAACAGGAATCTA 2968

## RESULT 2

US-08-972-631-3  
Sequence 3, Application US/08972631  
Patent No. 5856133  
GENERAL INFORMATION:  
APPLICANT: Stephens, Len  
TITLE OF INVENTION: G-BETA-GAMMA REGULATED  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,631  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/672,211  
FILING DATE: 27-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,277  
REFERENCE/DOCKET NUMBER: 8549-0005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)854-3660  
TELEFAX: (415)854-3694  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3808 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-972-631-3

## Query Match

Best Local Similarity 3.7%; Score 18; DB 2; Length 3808;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 TCAGAAACAGGAATCTA 428  
DB 2951 TCAGAAACAGGAATCTA 2968

## RESULT 3

US-08-972-629-3  
Sequence 3, Application US/08972629  
Patent No. 5859201  
GENERAL INFORMATION:  
APPLICANT: Stephens, Len  
TITLE OF INVENTION: G-BETA-GAMMA REGULATED  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park

STATE: California  
COUNTRY: USA  
ZIP: 94025

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,629  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/672,211  
FILING DATE: 27-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,277  
REFERENCE/DOCKET NUMBER: 8549-0005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)854-3660  
TELEFAX: (415)854-3694  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3808 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-972-629-3

## Query Match

Best Local Similarity 3.7%; Score 18; DB 2; Length 3808;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 411 TCAGAAACAGGAATCTA 428  
DB 2951 TCAGAAACAGGAATCTA 2968

## RESULT 4

US-08-972-630-3  
Sequence 3, Application US/08972630  
Patent No. 5869271  
GENERAL INFORMATION:  
APPLICANT: Stephens, Len  
TITLE OF INVENTION: G-BETA-GAMMA REGULATED  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,630  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/672,211  
FILING DATE: 27-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,277  
REFERENCE/DOCKET NUMBER: 8549-0005-999



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)854-3660  
TELEFAX: (415)854-3694  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3808 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-972-630-3

Query Match 3.7%; Score 18; DB 2; Length 3808;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 TCAGAAACAGGAATCTA 428  
|||||  
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 5  
US-08-672-211-3  
Sequence 3, Application US/08672211  
Patent No. 5874273

GENERAL INFORMATION:  
APPLICANT: Stephens, Len  
APPLICANT: Hawkins, Phillip T.  
TITLE OF INVENTION: G-BETA-GAMMA REGULATED  
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/672,211  
FILING DATE: 27-JUN-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,277  
REFERENCE/DOCKET NUMBER: 8549-0005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)854-3660  
TELEFAX: (415)854-3694  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3808 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-672-211-3

Query Match 3.7%; Score 18; DB 2; Length 3808;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 TCAGAAACAGGAATCTA 428  
|||||  
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 6

US-09-225-170-3  
Sequence 3, Application US/09225170  
Patent No. 6017763

GENERAL INFORMATION:  
APPLICANT: Stephens, Len  
APPLICANT: Hawkins, Phillip Thomas  
APPLICANT: Braselmann, Sylvia  
TITLE OF INVENTION: G-BETA-GAMMA REGULATED  
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds, LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2811

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/225,170  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/916,917  
FILING DATE: 15-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Abrams, Samuel B.  
REGISTRATION NUMBER: 30,605  
REFERENCE/DOCKET NUMBER: 8549-0006-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-493-4935  
TELEFAX: 650-493-5556  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3808 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: Linear  
US-09-225-170-3

Query Match 3.7%; Score 18; DB 3; Length 3808;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 TCAGAAACAGGAATCTA 428  
|||||  
DB 2951 TCAGAAACAGGAATCTA 2968

RESULT 7  
US-09-103-840A-1/C  
Sequence 1, Application US/09103840A  
Patent No. 6294328

GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA

ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

## Query Match

Best Local Similarity 3.7%; Score 18; DB 4; Length 4411529;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 CAGGAACGCCCTCAGCAG 372  
DB 2210673 CAGGAACGCCCTCAGCAG 2210656

## RESULT 8

US-09-541-941B-26/c  
Sequence 26, Application US/09541941B  
Patent No. 6309840  
GENERAL INFORMATION:  
APPLICANT: Wang, Jun  
APPLICANT: Shaw, Pang Chui  
APPLICANT: Paul, Pui-Hay  
APPLICANT: Ngan, But-and Pui Ngor Karenda Ngan  
TITLE OF INVENTION: POLYMERASE CHAIN REACTION - RESTRICTION FRAGMENT LENGTH POLYMORPHISM  
FILE REFERENCE: 2913/52188-ZA  
CURRENT APPLICATION NUMBER: US/09/541,941B  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 26  
LENGTH: 754  
TYPE: DNA  
ORGANISM: Tulipa Edulis  
US-09-541-941B-26

## Query Match

Best Local Similarity 3.5%; Score 17; DB 4; Length 754;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 GCGTCCCTCAGCCGGA 452  
DB 413 GCGTCCCTCAGCCGGA 397

## RESULT 9

US-08-553-888A-1/c  
Sequence 1, Application US/08553888A  
Patent No. 5723293  
GENERAL INFORMATION:  
APPLICANT: Huang  
TITLE OF INVENTION: DIAGNOSTIC METHOD AND KIT FOR DETERMINING RH BLOOD GROUP GENOTYPE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann & Baron, LLP  
STREET: 350 Jericho Turnpike  
CITY: Jericho  
STATE: New York  
COUNTRY: USA  
ZIP: 11753  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/553,888A  
FILING DATE: 11/06/95  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Dea, Sean W.  
REGISTRATION NUMBER: 37690  
REFERENCE/DOCKET NUMBER: 454-5

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 822-3550  
TELEFAX: (516) 822-3582  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1384 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-553-888A-1

## Query Match

Best Local Similarity 3.5%; Score 17; DB 1; Length 1384;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GCTCTGATCATATATC 39  
DB 635 GCTCTGATCATATATC 619

## RESULT 10

US-08-553-888A-2/c  
Sequence 2, Application US/08553888A  
Patent No. 5723293  
GENERAL INFORMATION:  
APPLICANT: Huang  
TITLE OF INVENTION: DIAGNOSTIC METHOD AND KIT FOR DETERMINING RH BLOOD GROUP GENOTYPE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann & Baron, LLP  
STREET: 350 Jericho Turnpike  
CITY: Jericho  
STATE: New York  
COUNTRY: USA  
ZIP: 11753  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/553,888A  
FILING DATE: 11/06/95  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Dea, Sean W.  
REGISTRATION NUMBER: 37690  
REFERENCE/DOCKET NUMBER: 454-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 822-3550  
TELEFAX: (516) 822-3582  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1466 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-553-888A-2

Query Match  
Best Local Similarity 3.5%; Score 17; DB 1; Length 1466;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GCTCTGATCATATATC 39  
DB 649 GCTCTGATCATATATC 633

## RESULT 11

US-09-032-742-1/c  
Sequence 1, Application US/09032742  
Patent No. 6255089

```

: GENERAL INFORMATION:
: APPLICANT: Tellier, Milt
: APPLICANT: Herrick-Davis, Katharine
: APPLICANT: Egan, Christina C.
: TITLE OF INVENTION: Constitutively Activated Serotonin
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Laurence Weinberger
: STREET: 882 S. Malack Street, Suite 103
: CITY: West Chester
: STATE: PA
: COUNTRY: USA
: ZIP: 19380-0053
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/032.742
: FILING DATE: 27-FEB-1998
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Weinberger, Laurence
: REGISTRATION NUMBER: 27,965
: REFERENCE/DOCKET NUMBER: 3086-4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (610) 431-1703
: TELEFAX: (610) 431-4181
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1566 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-032-742-1

Query Match      3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 345 TTACAACAACAGGAA 360
Db 1062 TTACAACAACAGGAA 1047

RESULT 12
US-09-032-742-12/c
: Sequence 12, Application US/09032742
: Patent No. 6255089
: GENERAL INFORMATION:
: APPLICANT: Tellier, Milt
: APPLICANT: Herrick-Davis, Katharine
: APPLICANT: Egan, Christina C.
: TITLE OF INVENTION: Constitutively Activated Serotonin
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Laurence Weinberger
: STREET: 882 S. Malack Street, Suite 103
: CITY: West Chester
: STATE: PA
: COUNTRY: USA
: ZIP: 19380-0053
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/032.742
: FILING DATE: 27-FEB-1998
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Weinberger, Laurence
: REGISTRATION NUMBER: 27,965
: REFERENCE/DOCKET NUMBER: 3086-4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (610) 431-1703
: TELEFAX: (610) 431-4181
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1566 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-032-742-12

Query Match      3.3%; Score 16; DB 4; Length 1566;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 345 TTACAACAACAGGAA 360
Db 1062 TTACAACAACAGGAA 1047

RESULT 13
US-09-032-742-13/c
: Sequence 13, Application US/09032742
: Patent No. 6255089
: GENERAL INFORMATION:
: APPLICANT: Tellier, Milt
: APPLICANT: Herrick-Davis, Katharine
: APPLICANT: Egan, Christina C.
: TITLE OF INVENTION: Constitutively Activated Serotonin
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Laurence Weinberger
: STREET: 882 S. Malack Street, Suite 103
: CITY: West Chester
: STATE: PA
: COUNTRY: USA
: ZIP: 19380-0053
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/032.742
: FILING DATE: 27-FEB-1998
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Weinberger, Laurence
: REGISTRATION NUMBER: 27,965
: REFERENCE/DOCKET NUMBER: 3086-4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (610) 431-1703
: TELEFAX: (610) 431-4181
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1566 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-032-742-13
```

Query Match  
Best Local Similarity 3.3%; Score 16; DB 4; Length 1566;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 345 TTACACAAACAGGAA 360  
|||||  
DB 1062 TTACACAAACAGGAA 1047

RESULT 14  
US-09-032-742-15/c  
Sequence 15, Application US/09032742  
Patent No. 6255089  
GENERAL INFORMATION:  
APPLICANT: Teitler, Milt  
APPLICANT: Herrick-Davis, Katharine  
TITLE OF INVENTION: Constitutively Activated Serotonin  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Laurence Weinberger  
STREET: 882 S. Mallick Street, Suite 103  
CITY: West Chester  
STATE: PA  
COUNTRY: USA  
ZIP: 19380-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,742  
FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Weinberger, Laurence  
REGISTRATION NUMBER: 27,965  
REFERENCE/DOCKET NUMBER: 3086-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 431-1703  
TELEFAX: (610) 431-4181  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1566 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-032-742-15

Query Match  
Best Local Similarity 3.3%; Score 16; DB 4; Length 1566;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 345 TTACACAAACAGGAA 360  
|||||  
DB 1062 TTACACAAACAGGAA 1047

RESULT 15  
US-09-032-742-16/c  
Sequence 16, Application US/09032742  
Patent No. 6255089  
GENERAL INFORMATION:  
APPLICANT: Teitler, Milt  
APPLICANT: Herrick-Davis, Katharine  
TITLE OF INVENTION: Constitutively Activated Serotonin  
NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Laurence Weinberger  
STREET: 882 S. Mallick Street, Suite 103  
CITY: West Chester  
STATE: PA  
COUNTRY: USA  
ZIP: 19380-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,742  
FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Weinberger, Laurence  
REGISTRATION NUMBER: 27,965  
REFERENCE/DOCKET NUMBER: 3086-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 431-1703  
TELEFAX: (610) 431-4181  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1566 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-032-742-16

Query Match  
Best Local Similarity 3.3%; Score 16; DB 4; Length 1566;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 345 TTACACAAACAGGAA 360  
|||||  
DB 1062 TTACACAAACAGGAA 1047

Search completed: October 27, 2002, 23:22:22  
Job time: 1738.87 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 17:46:54 ; Search time 122.349 Seconds  
(without alignments)  
6777.910 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483  
Sequence: 1 atgcgtgttgccctgcagg.....agtgattagcaccgaaaa 483

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

word size : 0

Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

1. N\_Genseq\_032802: \*

2. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1980.DAT: \*

3. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1981.DAT: \*

4. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1982.DAT: \*

5. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1983.DAT: \*

6. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1984.DAT: \*

7. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1985.DAT: \*

8. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1986.DAT: \*

9. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1987.DAT: \*

10. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1988.DAT: \*

11. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1989.DAT: \*

12. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1990.DAT: \*

13. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1991.DAT: \*

14. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1992.DAT: \*

15. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1993.DAT: \*

16. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1994.DAT: \*

17. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1995.DAT: \*

18. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1996.DAT: \*

19. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1997.DAT: \*

20. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1998.DAT: \*

21. /SIDS1/gcgdata/genseq/genseqn-emb1/NA1999.DAT: \*

22. /SIDS1/gcgdata/genseq/genseqn-emb1/NA2000.DAT: \*

23. /SIDS1/gcgdata/genseq/genseqn-emb1/NA2001.DAT: \*

24. /SIDS1/gcgdata/genseq/genseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	length	DB	ID	Description
1	483	100.0	483	22	AAE86247		DNA sequence of E.
2	483	100.0	768	22	AAE86248		DNA sequence of C.
3	118	24.4	118	22	AAE86254		PCR primer #4 used
4	104	21.5	110	22	AAE86252		PCR primer #2 used
5	102	21.1	102	22	AAE86255		PCR primer #5 used
6	75	15.5	94	22	AAE86253		PCR primer #3 used
7	74	15.3	111	22	AAE86251		PCR primer #1 used
8	40	8.3	110	22	AAE86256		PCR primer #6 used
9	20	4.1	3833	23	ABJ02577		Drosophila melanog

C	10	19	3	9	2562	22	AA159154	Human polynucleotide
C	11	18	3	7	1029	22	AAH52010	Mycobacterium tuberculosis
C	12	18	3	7	1300	18	AAAT66951	Asialoglycoprotein
C	13	18	3	7	1356	23	AAAS59780	Prophionibacterium
C	14	18	3	7	3684	23	AAAS84339	DNA encoding novel
C	15	18	3	7	3808	19	AAV046134	Porcine phospholipid
C	16	18	3	7	3808	20	AAV074100	Porcine G-protein
C	17	18	3	7	3808	21	AAZ96831	Pig p120 regulator
C	18	18	3	7	4890	23	AAAS68250	DNA encoding novel
C	19	18	3	7	11330	23	AAAS59544	Prophionibacterium
C	20	18	3	7	11736	22	AAAT22652	Human CDNA encoding
C	21	18	3	7	61313	23	AAAS92545	Prophionibacterium
C	22	18	3	7	4411529	22	AAAI99682	Mycobacterium tuberculosis
C	23	17	3	5	47	21	AAZ68405	Human map-related
C	24	17	3	5	123	22	ABAA48015	Human breast cell
C	25	17	3	5	123	22	ABAA65900	Human foetal liver
C	26	17	3	5	123	22	ABAA32986	Probe #11452 for g
C	27	17	3	5	123	22	AAAK14315	Human brain expres
C	28	17	3	5	123	22	AAKA40045	Human bone marrow
C	29	17	3	5	123	22	AAI20829	Probe #10762 for g
C	30	17	3	5	123	22	AAIA46064	Probe #14750 used t
C	31	17	3	5	123	22	AAI06540	Probe #6531 used t
C	32	17	3	5	401	22	AAAS56520	Human CDNA for an
C	33	17	3	5	436	22	AAAF67571	Novel human polynu
C	34	17	3	5	486	22	ABAA42886	Human breast cell
C	35	17	3	5	486	22	ABAA53307	Human foetal liver
C	36	17	3	5	486	22	ABAA23081	Probe #1547 for ge
C	37	17	3	5	486	22	AAAK01568	Human brain expres
C	38	17	3	5	486	22	AAK27012	Human bone marrow
C	39	17	3	5	486	22	AAI11615	Probe #1548 for ge
C	40	17	3	5	486	22	AAI13911	Probe #1597 used t
C	41	17	3	5	486	22	AAI015442	Probe #1533 used t
C	42	17	3	5	754	24	AAAD23833	Tulipa edulis rts1
C	43	17	3	5	808	19	AAAV28698	Ripening banana pu
C	44	17	3	5	808	22	AAH32197	Human olfactory re
C	45	17	3	5	881	22	AAH32196	Human olfactory re

## ALIGNMENTS

XX	AAAF86247				
XX	AAF86247	standard; DNA; 483 BP.			
XX	AAF86247;				
XX	11-JUL-2001	(first entry)			
DE	DNA sequence of E. coli optimised ospA gene 17E2.				
KW	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;				
KM	vaccine: OspA; salmonid rickettsial septicemia; rickettsial disease;				
KX	SRS; 17E2; ds.				
XX	Piscirickettsia salmonis.				
OS	Synthetic.				
XX					
FH	Key	Location/Qualifiers			
FT	CDS	1..483			
FT		/+tag= a			
FT		/partial			
FT		/product= "OspA"			
FT		/note= "Genus specific 17kDa antigen, the sequence does not include a stop codon"			
XX					
PN	CA2281913-A1.				
PD	17-MAR-2001.				
XX					
PP	17-SEP-1999;	99CA-2281913.			
PR	17-SEP-1999;	99CA-2281913.			

Best Local Similarity	100.0%;	Score 483;	DB 22;	Length 483;
Matches 483;	Conservative	Pred. No. 7,1e-243.		

[illegible]

## RESULT 2

ID	AAFP6248	standard; DNA: 768 BP.
AC	AAFP6248;	
DT	11-JUL-2001	(first entry)
DE	DNA sequence of c17e2 ospa construct with N-terminal fusion partner.	
KW	Polkilohermic fish; Piscirickettsia salmonis; rickettsial pathogen; vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;	
KW	SRS; 17E2; fusion construct; ds.	
OS	Piscirickettsia salmonis.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..768
FT		/+tag- a
FT		/partial
FT		/product= "C17E2 Ospa with N-terminal fusion partner"
FT		/note= "No stop codon is given"
FT		1..285
FT		/tag= b
FT	misc_feature	/note= "DNA encoding undefined N-terminal fusion partner"
FT		286..768
FT		/tag= c
FT		/note= "Optimised Ospa construct c17e2"
XX		
PN	CA2281913-A1.	
XX		
PD	17-MAR-2001.	
XX		
PF	17-SEP-1999;	99CA-2281913.
XX		
PR	17-SEP-1999;	99CA-2281913.
XX		
PA	(KAYW) KAY W W.	
PA	(BURI) BURIAN J.	
PA	(KUZ) KUZIK M A.	
XX		
PI	Kay WW, Burian J, Kuzik MA;	
XX		
DR	WPI: 2001-316844/34.	
DR	P-PSDB; AAB81128.	
XX		
PT	Method for protecting polkilohermic fish against salmonid rickettsial septicemia and other rickettsial diseases comprises administering a vaccine containing the Ospa protein of Piscirickettsia salmonis	
PT	Example 4; Fig 5; 35pp; English.	
CC	This invention relates to a method for the protection against infection of a polkilohermic fish by the bacterial pathogen, Piscirickettsia salmonis. The method comprises administering an immunogenic amount of P. salmonis specific antigen termed Ospa, or an immunogenic fragment of Ospa in the form of a vaccine. The method is used for protecting animals, particularly polkilohermic fish, against the bacterial pathogen P. salmonis. The method is also useful for protecting against salmonid rickettsial septicemia (SRS) and other rickettsial diseases. The sequence represents P. salmonis Ospa DNA termed C17E2 optimised for expression in Escherichia coli fused to DNA encoding an undefined N-terminal fusion partner. The protein encoded by this fusion construct is used in a vaccine to create an anti-Ospa antibody response.	
CC	Sequence 768 BP: 220 A..193 C..	

Query Match	100.00	99.99	99.98	99.97	99.96	99.95	99.94	99.93	99.92	99.91	99.90	99.89	99.88	99.87	99.86	99.85	99.84	99.83	99.82	99.81	99.80	99.79	99.78	99.77	99.76	99.75	99.74	99.73	99.72	99.71	99.70	99.69	99.68	99.67	99.66	99.65	99.64	99.63	99.62	99.61	99.60	99.59	99.58	99.57	99.56	99.55	99.54	99.53	99.52	99.51	99.50	99.49	99.48	99.47	99.46	99.45	99.44	99.43	99.42	99.41	99.40	99.39	99.38	99.37	99.36	99.35	99.34	99.33	99.32	99.31	99.30	99.29	99.28	99.27	99.26	99.25	99.24	99.23	99.22	99.21	99.20	99.19	99.18	99.17	99.16	99.15	99.14	99.13	99.12	99.11	99.10	99.09	99.08	99.07	99.06	99.05	99.04	99.03	99.02	99.01	99.00	98.99	98.98	98.97	98.96	98.95	98.94	98.93	98.92	98.91	98.90	98.89	98.88	98.87	98.86	98.85	98.84	98.83	98.82	98.81	98.80	98.79	98.78	98.77	98.76	98.75	98.74	98.73	98.72	98.71	98.70	98.69	98.68	98.67	98.66	98.65	98.64	98.63	98.62	98.61	98.60	98.59	98.58	98.57	98.56	98.55	98.54	98.53	98.52	98.51	98.50	98.49	98.48	98.47	98.46	98.45	98.44	98.43	98.42	98.41	98.40	98.39	98.38	98.37	98.36	98.35	98.34	98.33	98.32	98.31	98.30	98.29	98.28	98.27	98.26	98.25	98.24	98.23	98.22	98.21	98.20	98.19	98.18	98.17	98.16	98.15	98.14	98.13	98.12	98.11	98.10	98.09	98.08	98.07	98.06	98.05	98.04	98.03	98.02	98.01	98.00	97.99	97.98	97.97	97.96	97.95	97.94	97.93	97.92	97.91	97.90	97.89	97.88	97.87	97.86	97.85	97.84	97.83	97.82	97.81	97.80	97.79	97.78	97.77	97.76	97.75	97.74	97.73	97.72	97.71	97.70	97.69	97.68	97.67	97.66	97.65	97.64	97.63	97.62	97.61	97.60	97.59	97.58	97.57	97.56	97.55	97.54	97.53	97.52	97.51	97.50	97.49	97.48	97.47	97.46	97.45	97.44	97.43	97.42	97.41	97.40	97.39	97.38	97.37	97.36	97.35	97.34	97.33	97.32	97.31	97.30	97.29	97.28	97.27	97.26	97.25	97.24	97.23	97.22	97.21	97.20	97.19	97.18	97.17	97.16	97.15	97.14	97.13	97.12	97.11	97.10	97.09	97.08	97.07	97.06	97.05	97.04	97.03	97.02	97.01	97.00	96.99	96.98	96.97	96.96	96.95	96.94	96.93	96.92	96.91	96.90	96.89	96.88	96.87	96.86	96.85	96.84	96.83	96.82	96.81	96.80	96.79	96.78	96.77	96.76	96.75	96.74	96.73	96.72	96.71	96.70	96.69	96.68	96.67	96.66	96.65	96.64	96.63	96.62	96.61	96.60	96.59	96.58	96.57	96.56	96.55	96.54	96.53	96.52	96.51	96.50	96.49	96.48	96.47	96.46	96.45	96.44	96.43	96.42	96.41	96.40	96.39	96.38	96.37	96.36	96.35	96.34	96.33	96.32	96.31	96.
-------------	--------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-----

Best Local Similarity	100.0%;	Score 483;	DB 22;	Length 768;
Matches 483; Conservative	100.0%;	Pred. No. 7,1e-243.		

1 ATGCGTGGTTGCCCTGCACCCCAAGC	indels 0;	Gaps 0;
------------------------------	-----------	---------

|||||  
|||GAGCAGCTCCATCATTTACTCTGTATTTCCTGGCGGTGC 60  
|||||

```

Db 286 ATGCGTGGTTCCTGCAGAGGACGCTCTGATCATATCTCTGTTTTCCTGTTGGTTC 345
Qy 61 GCCCAGAACTTCAGCCGCCAGAAAGTTGGGCGGCCACCGGTGGGTTGGGGGGT 120
Db 346 GCCCAGAACTTCAGCCGCCAGAAAGTTGGGCGGCCACCGGTGGGTTGGGGGGT 405
Qy 121 GCCGCGCAGCTGTTCGTTAAAGGCTGTGTCGTGTGTCATGATGCGCATCGCGTGGCGTT 180
Db 406 GCCGCGCAGCTGTTCGTTAAAGGCTGTGTCGTGTGTCATGATGCGCATCGCGTGGCGTT 465
Qy 181 CTGGGCGCTGTGATGGCTCTAAATGCTGATGATGATGATGATGATGATGATGATGATGAT 240
Db 466 CTGGGCGCTGTGATGGCTCTAAATGCTGATGATGATGATGATGATGATGATGATGATGAT 525
Qy 241 CTGATACCACTCTCTGAAAAAGTGAAGCCGCCAGGTTACTGTTGGCGTAAATCCGGAC 300
Db 526 CTGATACCACTCTCTGAAAAAGTGAAGCCGCCAGGTTACTGTTGGCGTAAATCCGGAC 585
Qy 301 ACCGCTACAGCTACTCTGTGAAACCGGTTCCGACCTACAGCGTTTACACAAACAGAA 360
Db 586 ACCGCTACAGCTACTCTGTGAAACCGGTTCCGACCTACAGCGTTTACACAAACAGAA 645
Qy 361 CGCGCTCAGCAGTACTGTCGGAATTTACAGAGAAAGCCATGATCCGAGGTCAGAAACAG 420
Db 646 CGCGCTCAGCAGTACTGTCGGAATTTACAGAGAAAGCCATGATCCGAGGTCAGAAACAG 705
Qy 421 GAAATCTACGCGACCGCGTCCCTCAGCCGAGTGGCGCTGCGAGGTTAGCAGCGAA 480
Db 706 GAAATCTACGCGACCGCGTCCCTCAGCCGAGTGGCGCTGCGAGGTTAGCAGCGAA 765
Qy 481 AAA 483
Db 766 AAA 768

```

## RESULT 3

AAAF86254/C  
ID AAAF86254 standard; DNA: 118 BP.

AC AAAF86254;

DT 11-JUL-2001 (first entry)

DE PCR primer #4 used in cloning an optimisation of OspA gene.

XX Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;  
SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURJ/) BURIAN J.

PA (KUZK/) KUZK M A.

PI Kay WW, Burian J, Kuzk MA;

DR WPI; 2001-316844/34.

PT Method for protecting polkiothermic fish against salmonid rickettsial  
PT septicaemia and other rickettsial diseases comprises administering a  
PT vaccine containing the OspA protein of Piscirickettsia salmonis -  
XX Example 3; Fig 4B; 35pp; English.

```

CC This invention relates to a method for the protection against infection
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of
CC OspA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkiothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents a PCR primer used in the cloning and optimisation of
CC the P. salmonis OspA gene. The OspA gene is used in the method of the
CC invention.
SQ Sequence 118 BP; 21 A; 32 C; 28 G; 37 T; 0 other;
Query Match 24.4%; Score 118; DB 22; Length 118;
Best Local Similarity 100.0%; Pred. No. 1e-51;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 208 GGTACAGCATGGACGACGATTAATCAACTGAAACGAGTCTGAAAAAGTGAAC 267
Db 118 GGTACAGCATGGACGACGATTAATCAACTGAAACGAGTCTGAAAAAGTGAAC 59
Qy 268 GCCGCGCAGGTTACTGTTGGCGTAAATCCGACACCGGTAACAGTACTCTGTGAAC 325
Db 58 GCCGCGCAGGTTACTGTTGGCGTAAATCCGACACCGGTAACAGTACTCTGTGAAC 1

```

## RESULT 4

AAAF86252  
ID AAAF86252 standard; DNA: 110 BP.

AC AAAF86252;

DT 11-JUL-2001 (first entry)

DE PCR primer #2 used in cloning an optimisation of OspA gene.

XX Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
KW vaccine; OspA; salmonid rickettsial septicaemia; rickettsial disease;  
SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURJ/) BURIAN J.

PA (KUZK/) KUZK M A.

PI Kay WW, Burian J, Kuzk MA;

DR WPI; 2001-316844/34.

PT Method for protecting polkiothermic fish against salmonid rickettsial  
PT septicaemia and other rickettsial diseases comprises administering a  
PT vaccine containing the OspA protein of Piscirickettsia salmonis -  
XX Example 3; Fig 4B; 35pp; English.

CC This invention relates to a method for the protection against infection  
CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC P. salmonis specific antigen termed OspA, or an immunogenic fragment of  
CC OspA in the form of a vaccine. The method is used for protecting animals,  
CC particularly polkiothermic fish, against the bacterial pathogen  
CC P. salmonis. The method is also useful for protecting against salmonid  
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of  
 CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the  
 XX invention.  
 SO Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

Query Match 21.5%; Score 104; DB 22; Length 110;

Best Local Similarity 100.0%; Pred. No. 2.2e-44;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

53 TGGGTTCGCCAGCACTTCCAGCCGCCAGCAAGTTGGCCGCCACCGGTGGTGG 112

1 TGGGTTCGCCAGCACTTCCAGCCGCCAGCAAGTTGGCCGCCACCGGTGGTGG 112

113 GCGGTGTTCGCCAGCACTTCCAGCCGCCAGCAAGTTGGCCGCCACCGGTGGTGG 60

61 GCGGTGTTCGCCAGCACTTCCAGCCGCCAGCAAGTTGGCCGCCACCGGTGGTGG 104

# RESULT 5

AAFB6255/c

ID AAFB6255 standard; DNA; 102 BP.

AC AAFB6255;

11-JUL-2001 (first entry)

PCR primer #5 used in cloning an optimisation of Ospa gene.

Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;

SRS; 17E2; fusion construct; PCR primer; ss.

Piscirickettsia salmonis.

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAWW/) KAY W W.

(BUR/) BURIAN J.

(KUZ/) KUZIK M A.

Kay WW, Burian J, Kuzik MA;

WPI: 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial

septicaemia and other rickettsial diseases comprises administering a

vaccine containing the Ospa protein of Piscirickettsia salmonis

Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection

of a poikilothermic fish by the bacterial pathogen, Piscirickettsia

salmonis. The method comprises administering an immunogenic amount of a

P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a

P. salmonis. The method is also useful for protecting against animals,

rickettsial septicemia (SRS) and other rickettsial diseases. The present

sequence represents a PCR primer used in the cloning and optimisation of

the P. salmonis Ospa gene. The Ospa gene is used in the method of the

invention.

Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

Query Match 21.1%; Score 102; DB 22; Length 102;

Best Local Similarity 100.0%; Pred. No. 2.5e-43;

Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

334 ACCATACGAGCTTACCAACAAAGCAAGCGCGTACAGCACTACTGCGGAAATTCAGCAG 393

102 ACCATACGAGCTTACCAACAAAGCAAGCGCGTACAGCACTACTGCGGAAATTCAGCAG 393

394 AAAGCCATGATCCAGAGTCAAGAAACAGAAATCTACGGCACC 435

42 AAAGCCATGATCCAGAGTCAAGAAACAGAAATCTACGGCACC 1

# RESULT 6

AAFB6253

ID AAFB6253 standard; DNA; 94 BP.

AC AAFB6253;

11-JUL-2001 (first entry)

PCR primer #3 used in cloning an optimisation of Ospa gene.

Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;

SRS; 17E2; fusion construct; PCR primer; ss.

Piscirickettsia salmonis.

CA2281913-A1.

17-MAR-2001.

17-SEP-1999; 99CA-2281913.

17-SEP-1999; 99CA-2281913.

(KAWW/) KAY W W.

(BUR/) BURIAN J.

(KUZ/) KUZIK M A.

Kay WW, Burian J, Kuzik MA;

WPI: 2001-316844/34.

Method for protecting poikilothermic fish against salmonid rickettsial

septicaemia and other rickettsial diseases comprises administering a

vaccine containing the Ospa protein of Piscirickettsia salmonis

Example 3; Fig 4B; 35pp; English.

This invention relates to a method for the protection against infection

of a poikilothermic fish by the bacterial pathogen, Piscirickettsia

salmonis. The method comprises administering an immunogenic amount of a

P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a

P. salmonis. The method is also useful for protecting against animals,

rickettsial septicemia (SRS) and other rickettsial diseases. The present

sequence represents a PCR primer used in the cloning and optimisation of

the P. salmonis Ospa gene. The Ospa gene is used in the method of the

invention.

Sequence 94 BP; 18 A; 20 C; 35 G; 21 T; 0 other;

Query Match 15.5%; Score 75; DB 22; Length 94;

Best Local Similarity 100.0%; Pred. No. 3.4e-29;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

158 CGATGCCATCGCGGTGGTTCGGCGGCTGATGGTCTTAAATCGTCAGACGA 217

20 CGATGCCATCGCGGTGGTTCGGCGGCTGATGGTCTTAAATCGTCAGACGA 217

218 TGGACCAAGAGGATA 232



Db 80 TGGACGACAGATA 94

# RESULT 7

ID AAF86251 standard; DNA; 111 BP.

XX AAF86251;

DT 11-JUL-2001 (first entry)

DE PCR primer #1 used in cloning an optimisation of Ospa gene.

KM Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine: OSPA; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer: ss.

OS Piscirickettsia salmonis.

PN CA281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZV/) KUZV M A.

PI Kay WM, Burian J, Kuzyk MA;

PS WPI; 2001-316844/34.

PT Method for protecting polkillothermic fish against salmonid rickettsial

PT septicemia and other rickettsial diseases comprises administering a

PT vaccine containing the Ospa protein of Piscirickettsia salmonis

XX Example 3; Fig 4B; 35pp; English.

XX This invention relates to a method for the protection against infection

CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia

CC salmonis. The method comprises administering an immunogenic amount of a

CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of

CC OSPA in the form of a vaccine. The method is used for protecting animals,

CC particularly polkillothermic fish, against the bacterial pathogen

CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of

CC the P. salmonis OSPA gene. The Ospa gene is used in the method of the

XX invention.

XX Sequence 111 BP; 16 A; 32 C; 30 G; 33 T; 0 other;

SQ

Query Match 15.3%; Score 74; DB 22; Length 111;

Best Local Similarity 100.0%; Pred. No. 1.1e-28; Indels 0; Gaps 0;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTTGCTGACGAGGAGCTCTGATCATTAATCTGTTTCTGTTGCTGTTGC 60

Db 38 ATGGGTGTTGCTGACGAGGAGCTCTGATCATTAATCTGTTTCTGTTGCTGTTGC 97

OY 61 GCCCAGAACTTCAG 74

Db 98 GCCCAGAACTTCAG 111

# RESULT 8

ID AAF86256/C

XX AAF86256 standard; DNA; 110 BP.

XX 11-JUL-2001 (first entry)

DE PCR primer #6 used in cloning an optimisation of Ospa gene.

KM Polkillothermic fish: Piscirickettsia salmonis; rickettsial pathogen;

KW vaccine: OSPA; salmonid rickettsial septicemia; rickettsial disease;

XX SRS; 17E2; fusion construct; PCR primer: ss.

OS Piscirickettsia salmonis.

PN CA281913-A1.

PD 17-MAR-2001.

PF 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PA (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZV/) KUZV M A.

PI Kay WM, Burian J, Kuzyk MA;

PS WPI; 2001-316844/34.

PT Method for protecting polkillothermic fish against salmonid rickettsial

PT septicemia and other rickettsial diseases comprises administering a

PT vaccine containing the Ospa protein of Piscirickettsia salmonis

XX Example 3; Fig 4B; 35pp; English.

XX This invention relates to a method for the protection against infection

CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia

CC salmonis. The method comprises administering an immunogenic amount of a

CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of

CC OSPA in the form of a vaccine. The method is used for protecting animals,

CC particularly polkillothermic fish, against the bacterial pathogen

CC P. salmonis. The method is also useful for protecting against salmonid

CC rickettsial septicemia (SRS) and other rickettsial diseases. The present

CC sequence represents a PCR primer used in the cloning and optimisation of

CC the P. salmonis OSPA gene. The Ospa gene is used in the method of the

XX invention.

XX Sequence 110 BP; 17 A; 34 C; 27 G; 32 T; 0 other;

SQ

Query Match 8.3%; Score 40; DB 22; Length 110;

Best Local Similarity 100.0%; Pred. No. 7.3e-11; Indels 0; Gaps 0;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 444 TCAGCCGATGCGCGCTGCGAGGTGATTAGCAGCAAAAA 483

Db 76 TCAGCCGATGCGCGCTGCGAGGTGATTAGCAGCAAAAA 37

# RESULT 9

ID ABL02577 standard; cDNA; 3833 BP.

XX ABL02577;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2213.

KM Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB58474.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Claim 1; SEQ ID NO 2213; 21pp + Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB16176-AB16175) and the encoded proteins  
 CC (AB16176-AB16175).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIFO  
 CC at ftp.wifo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 3833 BP; 1094 A; 1030 C; 858 G; 851 T; 0 other;  
 Query Match  
 Best Local Similarity 4.1%; Score 20; DB 23; Length 3833;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 159 GATGGCCATGCGCGCGCG 178  
 DB 1428 GATGGCCATGCGCGCGCG 1447  
 ID AAI59154  
 XX AAI59154 standard; cDNA; 2562 BP.  
 AC AAI59154;  
 XX  
 DT 22-OCT-2001 (first entry)  
 DE Human polynucleotide SEQ ID NO 1357.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoketic;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 PD 26-JUL-2001.  
 PD 26-JUL-2001.  
 PF 26-DEC-2000; 2000WO-US34263.  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehman T, Xu C, Xue A, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 DR WPI; 2001-442253/47.  
 DR P-PSDB; AAM39998.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PS Claim 1; SEQ ID NO 1357; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localized neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 2562 BP; 524 A; 794 C; 791 G; 450 T; 3 other;  
 Query Match  
 Best Local Similarity 3.9%; Score 19; DB 22; Length 2562;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 259 AAGGTGAAGCCGCCAGG 277  
 DB 1197 AAGGTGAAGCCGCCAGG 1215  
 ID AAI52010/c  
 XX AAI52010 standard; DNA; 1029 BP.  
 AC AAI52010;  
 XX  
 DT 04-SEP-2001 (first entry)  
 DE Mycobacterium tuberculosis potential drug target gene SEQ ID 64.  
 XX  
 KW Mycobacterium tuberculosis potential drug target gene SEQ ID 64.  
 KW Drug target; growth; organism viability; characterisation; ds.  
 KW Mycobacterium tuberculosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO200135317-A1.  
 PD 17-MAY-2001.  
 PD 17-MAY-2001.  
 PF 13-NOV-2000; 2000WO-US31152.  
 PR 12-NOV-1999; 99US-0165086.  
 PR 12-NOV-1999; 99US-0165124.  
 PR 01-FEB-2000; 2000US-0179531.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Eisenberg D, Rotstein SH, Marcotte EM;

```

XX  WPI: 2001-329193/34.
DR  P-PSDB: AAG81159.
XX
PT  Identifying nucleotide or polypeptide sequence for use as drug target,
PT  involves providing algorithm that analyzes a functional relationship
PT  between nucleotide or polypeptide sequences, and comparing the
PT  sequences -
XX
PS  Disclosure: Page 91; 207pp: English.
XX
CC  This invention relates to a method for identifying a nucleotide or
CC  polypeptide sequence that may be a drug target, or essential for growth
CC  or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC  represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC  tuberculosis proteins which are potential drug targets. The DNA and
CC  protein sequences are used to illustrate the method of the invention. The
CC  method involves providing an unknown nucleotide or polypeptide sequences,
CC  and comparing it to a number of sequences along with at least one
CC  algorithm capable of analysing a functional relationship between
CC  nucleotide and polypeptide sequences. The method is useful for
CC  characterising the function of nucleic acids and polypeptides that may be
CC  useful as a target for a drug or essential for the growth or viability of
CC  an organism.
XX
SQ  Sequence 1029 BP; 169 A; 338 C; 341 G; 181 T; 0 other:
Query Match          3.7%; Score 18; DB 22; Length 1029;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  355 CAGGAACGCCGTCAGCAG 372
    |||||||||||||||
DB  75 CAGGAACGCCGTCAGCAG 58

RESULT 12
AAT66951
ID  AAT66951 standard; cDNA; 1300 BP.
XX
AC  AAT66951;
XX
XX  23-NOV-1997 (first entry)
XX
DE  Asialoglycoprotein receptor L-H2 cDNA.
XX
XX  Asialoglycoprotein receptor L-H2; AGPR; autoimmune hepatitis;
XX  autoantibody; diagnosis; liver; ds.
XX
OS  Homo sapiens.
XX
FH  Key          Location/Qualifiers
FT  CDS          191..1054
    /tag=a
XX
XX  EP773289-A2.
XX
PD  14-MAY-1997.
XX
XX  20-AUG-1996; 96EP-0113349.
XX
XX  21-AUG-1995; 95JP-0212118.
XX
PA  (TOFU ) TONEN CORP.
XX
PI  Hasegawa A, Kiyosawa K, Ohue C, Tanida E, Yagi S;
PI  Yano A;
XX
XX  WPI: 1997-261316/24.
XX  P-PSDB: AAM15246.
XX
PT  Asialoglycoprotein receptor H1 and L-H2 soluble derivatives -
PT  comprise extracellular domains, optionally also with cytoplasmic

```

```

PT  domains, useful for autoimmune hepatitis diagnosis
XX
XX  Example 2; Page 35-34; 40pp: English.
PS
XX  This cDNA clone codes for human asialoglycoprotein receptor (AGPR)
XX  L-H2 (AAM15246), a membrane protein expressed specifically by
XX  hepatocytes. The appearance of autoantibodies against AGPR can be
XX  used as an indicator for autoimmune hepatitis (AIH). The cDNA
XX  clone for AGPR L-H2 was isolated from human liver cDNA by PCR
XX  amplification (see AAT66956-61). Subregions coding for the
XX  extracellular domain, optionally also with the cytoplasmic domain,
XX  can be amplified by PCR (see AAT66966-69), cloned into appropriate
XX  vectors and used for large-scale prodn. of AGPR derivatives (see
XX  AAM15251-52), free of contaminating liver antigens, in bacterial
XX  (pref. E. coli) or animal (pref. mammalian) host cells. Such AGPR
XX  L-H2 derivatives can be used in a claimed method for detecting or
XX  measuring anti-AGPR antibody. AGPR H12 cDNA has also been cloned
XX  (see AAT66950).
SQ  Sequence 1300 BP; 302 A; 395 C; 350 G; 253 T; 0 other:
Query Match          3.7%; Score 18; DB 18; Length 1300;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  445 CAGCCGATGCCGCTGG 462
    |||||||||||||||
DB  962 CAGCCGATGCCGCTGG 979

RESULT 13
AAS59780/c
ID  AAS59780 standard; DNA; 1356 BP.
XX
XX  AAS59780;
XX
XX  13-FEB-2002 (first entry)
XX
DE  Propionibacterium acnes immunogenic protein encoding DNA #275.
XX
XX  SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX  uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX  inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX  dermatological; osteopathic; neuroprotectant; ds.
XX
OS  Propionibacterium acnes.
XX
XX  WO200181581-A2.
XX
XX  01-NOV-2001.
XX
XX  20-APR-2001; 2001WO-US12865.
XX
XX  21-APR-2000; 2000US-199047P.
XX  02-JUN-2000; 2000US-208841P.
XX  07-JUL-2000; 2000US-216747P.
XX
PA  (CORI-) CORIXA CORP.
XX
XX  Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX  L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX  WPI: 2001-616774/71.
XX
XX  Propionibacterium acnes polypeptides and nucleic acids useful for
XX  PT  vaccinating against and diagnosing infections, especially useful for
XX  PT  treating acne vulgaris -
XX
PS  Claim 1; SEQ ID No 275; 1069pp; English.
XX
XX  Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX  CC  Propionibacterium acnes immunogenic polypeptides. The proteins and their
XX  CC  associated DNA sequences are used in the treatment, prevention and

```

CC diagnosis of medical conditions caused by P. acnes. The disorders include  
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and  
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
 CC in infections of bone, joints and the central nervous system, however it  
 CC is particularly involved in the inflammatory lesions associated with acne  
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
 CC patient comprises contacting a sample with a binding agent that binds to  
 CC the proteins of the invention and determining the amount of bound protein  
 CC in the sample. The polypeptides may be used as antigens in the production  
 CC of antibodies specific for P. acnes proteins. These antibodies can be  
 CC used to downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 1356 BP; 267 A; 395 C; 429 G; 264 T; 1 other;

## Query Match

Best Local Similarity 3.7%; Score 18; DB 23; Length 1356;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 GCCATCGCGCGTGGCGTT 180  
 ||||||||||||||||  
 DB 426 GCCATCGCGCGTGGCGTT 409

## RESULT 14

AAS84339  
 ID AAS84339 standard; cDNA; 3684 BP.

AC AAS84339;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #20143.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dirmannac RT, Liu C, Tang YF;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG20152.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PI Claim 1; SEQ ID NO 20143; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful for medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 3684 BP; 851 A; 981 C; 1107 G; 743 T; 2 other;

## Query Match

Best Local Similarity 3.7%; Score 18; DB 23; Length 3684;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 TGCAGGCGACCTCTCTGA 31  
 ||||||||||||||||  
 DB 617 TGCAGGCGACCTCTCTGA 634

## RESULT 15

AAV04634  
 ID AAV04634 standard; cDNA; 3808 BP.

AC AAV04634;

DT 17-AUG-1998 (first entry)

DE Porcine phosphoinositide 3OH-kinase p120 subunit cDNA.

KM G-beta-gamma regulated phosphatidylinositol-3' kinase; p1g;  
 KM phosphoinositide 3OH-kinase; PI3K; signal transduction;  
 KM phosphatidylinositol (3,4,5)-triphosphate; G-protein; receptor;  
 KM transgenic animal; knockout animal; inflammation; arthritis;  
 KM septic shock; adult respiratory distress syndrome; pneumonia;  
 KM asthma; allergy; reperfusion injury; atherosclerosis; cancer;  
 KM Alzheimer's disease; cancer; antitense; ribozyme; diagnosis;  
 KM therapy; drug screening; ss.

OS Sus scrofa.

PN WO9749818-A2.

PD 31-DEC-1997.

PF 26-JUN-1997; 97WO-US11219.

PR 27-JUN-1996; 96US-0672211.

PR (ONX-) ONYX PHARM.

PA Baselmann S, Hawkins PT, Stephens L;

DR WPI; 1998-077181/07.

DR P-PSDB; AAW23948.

PT DNA encoding G-beta-gamma regulated phosphatidylinositol-3' kinase,  
 PT p101 and p120 subunits - useful for diagnosis, drug screening,  
 PT clinical trial monitoring and treatment of inflammatory disorders

PI Claim 54; Fig 3A-C; 151pp; English.

CC

```
XX
CC This cDNA sequence codes for the p120 catalytic subunit (see
CC AAW23948) of pig G-protein regulated phosphatidylinositol-3' kinase,
CC a heterodimeric enzyme which produces the intracellular messenger
CC phosphatidylinositol (3',4',5)-triphosphate in response to activation
CC of trimeric G protein-linked receptors. This novel protein, which
CC also contains a regulatory subunit, p101 (see AAV04633), is found in
CC cells of haematopoietic origin and is involved in immune system
CC responses which cause inflammation. p120 cDNA was obtained from a
CC pig neutrophil cDNA library using a degenerate oligonucleotide
CC probe (see AAT9713) based on an isolated p120 tryptic peptide. The
CC p120 cDNA clone in pCMV3mycpi20 is deposited as ATCC 97637. The
CC invention encompasses pig and human p101 and p120 nucleotides,
CC host cell expression systems, p101 and p120 proteins (see
CC AAW23946-49), fusion proteins, polypeptides and peptides, antibodies
CC to these proteins, and transgenic animals and knockout animals.
CC Compounds which are useful for treating inflammatory response
CC disorders can be identified by screening assays using a G protein
CC activated PI3K, or a cultured host cell that expresses the p101
CC gene. Antagonists of G protein stimulated PI3K (acting through the
CC p101 subunit, especially by disrupting the interaction between the
CC p101 and p120 subunits) can be used to treat arthritis, septic
CC shock, adult respiratory distress syndrome (ARDS), pneumonia,
CC asthma, allergies, reperfusion injury, atherosclerosis, cancer and
CC Alzheimer's disease. The nucleic acids and their products can also
CC be used for diagnosis, drug screening and clinical trial monitoring
CC of inflammatory diseases.
CC
XX
SQ Sequence 3808 BP; 1010 A; 965 C; 953 G; 880 T; 0 other;
Query Match 3.7%; Score 18; DB 19; Length 3808;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 411 TCAGAAACAGGAATCTA 428
| | | | | | | | | | | | | | | |
Db 2951 TCAGAAACAGGAATCTA 2968
```

Search completed: October 27, 2002, 20:42:30  
Job time : 130.349 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 18:35:09 : Search time 934.579 Seconds  
(without alignments)  
10815.053 Million cell updates/sec

Title: US-09-677-374-3

Perfect score: 483  
Sequence: 1 atgcgtggtgtgcctgcaggg.....aggtgattagcaccgaaaaa 483

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 segs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenBank: 1: gb\_ba: 2: gb\_hgt: 3: gb\_in: 4: gb\_ov: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: gb\_vl: 16: em\_ba: 17: em\_fun: 18: em\_hum: 19: em\_mu: 20: em\_mu: 21: em\_mu: 22: em\_mu: 23: em\_mu: 24: em\_mu: 25: em\_mu: 26: em\_mu: 27: em\_mu: 28: em\_mu: 29: em\_mu: 30: em\_mu: 31: em\_mu: 32: em\_mu: 33: em\_mu:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result Query Match Length DB ID Description

C	1	22	4.6	589	8	HJRNA1956	AJ001956	Hylomecon
C	2	22	4.6	594	8	CMRNA1959	AJ001959	Chellidoni
C	3	22	4.6	596	8	CMRNA1958	AJ001958	Chellidoni
C	4	22	4.6	596	8	CMRNA1960	AJ001960	Chellidoni
C	5	22	4.6	596	8	CMRNA1960	AJ001960	Chellidoni
C	6	22	4.6	7286	8	AE005332	AE005332	Zizania t
C	7	22	4.6	10029	1	AE005325	AE005325	Escherich
C	8	22	4.6	11809	1	AE005290	AE005290	Escherich
C	9	22	4.6	13505	1	AE005441	AE005441	Escherich
C	10	22	4.6	13624	1	AE006459	AE006459	Escherich
C	11	22	4.6	103608	9	AL139161	AL139161	Human DNA
C	12	22	4.6	222605	1	AP002555	AP002555	Escherich
C	13	22	4.6	286485	1	AP002557	AP002557	Escherich
C	14	22	4.6	291136	1	AP002556	AP002556	Escherich
C	15	21	4.3	702	8	AGHGBEHA	L13954	Agoseris he
C	16	21	4.3	705	8	MIRGBEHA	L13955	Microseris
C	17	21	4.3	706	8	KGIRGBEH	L20484	Ritiga bifl
C	18	21	4.3	706	8	KGIRGBEH	L13948	Ritiga bifl
C	19	21	4.3	706	8	LAURGBEHA	L13957	Lactuca sat
C	20	21	4.3	708	8	AF422126	AF422126	Microseri
C	21	21	4.3	748	8	MC069708	U69708	Microseris
C	22	21	4.3	750	8	AF386496	AF386496	Microseri
C	23	21	4.3	750	8	AF386497	AF386497	Microseri
C	24	21	4.3	750	8	AF386498	AF386498	Microseri
C	25	21	4.3	750	8	AF386499	AF386499	Microseri
C	26	21	4.3	750	8	AF386500	AF386500	Microseri
C	27	21	4.3	750	8	AF386501	AF386501	Microseri
C	28	21	4.3	750	8	AF386502	AF386502	Microseri
C	29	21	4.3	750	8	MAU69709	U69709	Microseri
C	30	21	4.3	750	8	MAU69709	U69709	Microseri
C	31	21	4.3	750	8	MBU69703	U69703	Microseris
C	32	21	4.3	750	8	MPU69705	U69705	Microseris
C	33	21	4.3	751	8	AF386495	AF386495	Utopappus
C	34	21	4.3	751	8	MEU69707	U69707	Microseris
C	35	21	4.3	752	8	MLU69704	U69704	Microseris
C	36	21	4.3	754	8	AF386493	AF386493	Agoseris
C	37	21	4.3	754	8	MDU69706	U69706	Microseris
C	38	21	4.3	774	8	AF262890	AF262890	Lepanthes
C	39	21	4.3	777	8	AF262892	AF262892	Lepanthes
C	40	21	4.3	779	8	AF262891	AF262891	Lepanthes
C	41	21	4.3	175144	2	AC068502	AC068502	Mus muscu
C	42	20	4.1	610	8	AF244734	AF244734	Triguera
C	43	20	4.1	616	8	AF201690	AF201690	Lawnsonia
C	44	20	4.1	619	8	AF229747	AF229747	Delariprea
C	45	20	4.1	643	8	AF465874	AF465874	Milleria

## ALIGNMENTS

RESULT 1  
HJRNA1956 589 bp DNA linear PLN 23-MAR-2000  
LOCUS HJRNA1956 Japonica 5.8S rRNA gene.  
DEFINITION HJRNA1956  
ACCESSION AJ001956  
VERSION AJ001956.1 GI:3392901  
KEYWORDS 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer; ITS1; ITS2.  
SOURCE Hylomecon japonica.  
ORGANISM Hylomecon japonica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales; Papaveraceae; Hylomecon.  
REFERENCE 1 (bases 1 to 589)  
Biatner, F.R. and Kaderelt, J.W.  
TITLE Morphological evolution and ecological diversification of the forest-dwelling poppies (Papaveraceae: Chelidonioidae) as deduced from a molecular phylogeny of the ITS region  
JOURNAL Plant Syst. Evol. 219, 181-197 (1999)  
REFERENCE 2 (bases 1 to 589)  
Biatner, F.R.  
TITLE Direct Submission  
JOURNAL Submitted (02-OCT-1997) Biatner F.R., Dept. of Taxonomy, Institute

for Plant Genetics and Crop Plant Research, Gatersleben, D-06466,  
GERMANY

FEATURES  
source Location/Qualifiers

misc\_feature  
1..589  
/organism="Hymenoclea japonica"  
/db\_xref="taxon:71217"

rRNA  
1..225  
/note="internal transcribed spacer 1"

gene  
226..387  
/gene="5.8S rRNA"  
/product="5.8S ribosomal RNA"

misc\_feature  
388..589  
/gene="5.8S rRNA"

BASE COUNT 125 a 169 c 166 g 129 t  
ORIGIN

Query Match 4.6%; Score 22; DB 8; Length 589;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGTGCCCTCAGCCGATGACC 457  
Db 370 GCGTGCCCTCAGCCGATGACC 349

RESULT 2  
CMRNA1959/c

LOCUS CMRNA1959 594 bp DNA linear PLN 23-MAR-2000  
DEFINITION Chelidonium majus 5.8S rRNA gene.  
ACCESSION AJ001959.1 GI:3393038  
VERSION 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer;  
KEYWORDS ITS1; ITS2.

SOURCE  
ORGANISM Chelidonium majus.  
Chelidonium majus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
Papaveraceae; Chelidonium.  
1 (bases 1 to 594)  
Blattner, F.R. and Kaderelt, J.W.  
Morphological evolution and ecological diversification of the  
forest-dwelling poppies (Papaveraceae; Chelidoniaceae) as deduced  
from a molecular phylogeny of the ITS region  
Plant Syst. Evol. 219, 181-197 (1999)  
2 (bases 1 to 594)  
Blattner, F.R.  
Direct Submission  
Submitted (02-OCT-1997) Blattner F.R., Dept. of Taxonomy, Institute  
for Plant Genetics and Crop Plant Research, Gatersleben, D-06466,  
GERMANY

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source Location/Qualifiers  
1..594  
/organism="Chelidonium majus"  
/strain="Mainz"  
/sub\_species="majus"  
/db\_xref="taxon:71251"

misc\_feature  
1..229  
/note="internal transcribed spacer 1"

rRNA  
230..391  
/gene="5.8S rRNA"  
/product="5.8S ribosomal RNA"

gene  
230..391  
/gene="5.8S rRNA"

misc\_feature  
392..594  
/note="internal transcribed spacer 2"

BASE COUNT 123 a 168 c 173 g 130 t  
ORIGIN

Query Match 4.6%; Score 22; DB 8; Length 594;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGTGCCCTCAGCCGATGACC 457  
Db 373 GCGTGCCCTCAGCCGATGACC 352

RESULT 3  
CMRNA1958/c

LOCUS CMRNA1958 596 bp DNA linear PLN 23-MAR-2000  
DEFINITION Chelidonium majus 5.8S rRNA gene.  
ACCESSION AJ001958.1 GI:3393037  
VERSION 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer;  
KEYWORDS ITS1; ITS2.

SOURCE  
ORGANISM Chelidonium majus.  
Chelidonium majus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
Papaveraceae; Chelidonium.  
1 (bases 1 to 596)  
Blattner, F.R. and Kaderelt, J.W.  
Morphological evolution and ecological diversification of the  
forest-dwelling poppies (Papaveraceae; Chelidoniaceae) as deduced  
from a molecular phylogeny of the ITS region  
Plant Syst. Evol. 219, 181-197 (1999)  
2 (bases 1 to 596)  
Blattner, F.R.  
Direct Submission  
Submitted (02-OCT-1997) Blattner F.R., Dept. of Taxonomy, Institute  
for Plant Genetics and Crop Plant Research, Gatersleben, D-06466,  
GERMANY

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source Location/Qualifiers  
1..596  
/organism="Chelidonium majus"  
/sub\_species="asiaticum Hara"  
/db\_xref="taxon:71251"

misc\_feature  
1..229  
/note="internal transcribed spacer 1"

rRNA  
230..391  
/gene="5.8S rRNA"

gene  
230..391  
/product="5.8S ribosomal RNA"

misc\_feature  
392..596  
/gene="5.8S rRNA"

BASE COUNT 124 a 169 c 171 g 132 t  
ORIGIN

Query Match 4.6%; Score 22; DB 8; Length 596;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGTGCCCTCAGCCGATGACC 457  
Db 375 GCGTGCCCTCAGCCGATGACC 354

RESULT 4  
CMRNA1960/c

LOCUS CMRNA1960 596 bp DNA linear PLN 23-MAR-2000  
DEFINITION Chelidonium majus 5.8S rRNA gene.  
ACCESSION AJ001960.1 GI:3393039  
VERSION 5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer;  
KEYWORDS ITS1; ITS2.

SOURCE  
ORGANISM Chelidonium majus.  
Chelidonium majus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
Papaveraceae; Chelidonium.  
1 (bases 1 to 596)  
Blattner, F.R. and Kaderelt, J.W.  
Morphological evolution and ecological diversification of the

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL



forest-dwelling poppies (Papaveraceae: Chelidonioidae) as deduced from a molecular phylogeny of the ITS region

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-OCT-1997) Blattner F.R., Dept. of Taxonomy, Institute for Plant Genetics and Crop Plant Research, Gatersleben, D-06466, GERMANY

FEATURES  
source  
Location/Qualifiers

1. .596  
/organism="Chelidonium majus"  
/strain="Neustadt/W."  
/sub-species="majus"  
/db\_xref="taxon:71251"

misc\_feature  
1. .228  
/note="Internal transcribed spacer 1"

rRNA  
229. .390  
/gene="5.8S rRNA"  
/product="5.8S ribosomal RNA"

gene  
391. .596  
/gene="5.8S rRNA"

misc\_feature  
/note="Internal transcribed spacer 2"

BASE COUNT 122 a 172 c 171 g 131 t  
ORIGIN  
Query Match 4.6%; Score 22; DB 8; Length 596;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGGCGCCCTCAGCCGATGGCC 457  
Db 373 GCGGCGCCCTCAGCCGATGGCC 352

RESULT 5  
AF169233/c 648 bp DNA linear PLN 21-NOV-1999  
LOCUS  
DEFINITION  
Zizania texana 18S ribosomal RNA gene, partial sequence; Internal transcribed spacer 1, 5.8S ribosomal RNA gene and Internal transcribed spacer 2, complete sequence; and 25S ribosomal RNA gene, partial sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AF169233.1 GI:6456574  
Zizania texana.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (14-JUL-1999) Biology, Southwest Texas State University, San Marcos, TX 78666, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (14-JUL-1999) Biology, Southwest Texas State University, San Marcos, TX 78666, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (14-JUL-1999) Biology, Southwest Texas State University, San Marcos, TX 78666, USA

FEATURES  
source  
Location/Qualifiers  
1. .648  
/organism="Zizania texana"  
/db\_xref="taxon:34180"  
/country="USA"  
misc\_RNA  
30. .237  
/product="18S ribosomal RNA"

rRNA  
238. .404  
/note="ITS1"  
/product="Internal transcribed spacer 1"  
rRNA  
/product="5.8S ribosomal RNA"

misc\_RNA

405. .606  
/note="ITS2"  
/product="Internal transcribed spacer 2"

BASE COUNT 136 a 211 c 192 g 109 t  
ORIGIN

Query Match 4.6%; Score 22; DB 8; Length 648;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCGGCGCCCTCAGCCGATGGCC 457  
Db 376 GCGGCGCCCTCAGCCGATGGCC 355

RESULT 6  
AE005332 7286 bp DNA linear BCT 21-MAR-2001  
LOCUS  
DEFINITION  
Escherichia coli O157:H7 EDL933 genome, contig 2 of 3, section 1 of 55.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AE005332.1 GI:12514830  
Escherichia coli O157:H7 EDL933.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE  
AUTHORS  
Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grodebeck, E.J., Davis, N.W., Lim, A., Diallanla, E., Potamocis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.  
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

1 (bases 1 to 7286)  
Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grodebeck, E.J., Davis, N.W., Lim, A., Diallanla, E., Potamocis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.  
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

2 (bases 1 to 7286)  
Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grodebeck, E.J., Davis, N.W., Lim, A., Diallanla, E., Potamocis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES  
source  
Location/Qualifiers

1. .7286  
/organism="Escherichia coli O157:H7 EDL933"  
/strain="EDL933"  
/serotype="O157:H7"  
/db\_xref="taxon:155864"  
/note="enterohemorrhagic"  
misc\_feature  
1. .7286  
/note="O-island #52; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"

gene  
316. .642  
/gene="21901"  
316. .642  
/gene="21901"

CDS  
316. .642  
/gene="21901"  
/function="orf; Other or unknown (Phage or Prophage Related)"  
/note="No significant matches"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"

/protein\_id="AAG55994.1"  
/db\_xref="GI:12514831"  
/translation="MTALLLEIKANLHVNDADDEMLMDKVRQATVLLVLYIGSSR"

gene  
CDS  
652..1002  
/gene="21902"  
/function="putative enzyme; DNA packaging, phage assembly (phage or prophage related)"  
/note="Residues 6 to 116 of 116 are 44.14 pct identical to 916901597|gb|AF31100.1| (AF069529) putative head-tail adaptor [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="putative head-tail adaptor of prophage CP-933X"  
/protein\_id="AAG55995.1"  
/db\_xref="GI:12514832"

gene  
CDS  
999..1244  
/gene="21903"  
/function="orf; Other or unknown (phage or prophage related)"  
/note="Residues 1 to 55 of 81 are 67.27 pct identical to 916901598|gb|AF31101.1| (AF069529) gp10 [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"  
/protein\_id="AAG55996.1"  
/db\_xref="GI:12514833"

gene  
CDS  
1180..1446  
/gene="21905"  
/function="orf; Other or unknown (phage or prophage related)"  
/note="Residues 1 to 88 of 88 are 73.86 pct identical to 916901598|gb|AF31101.1| (AF069529) gp10 [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"  
/protein\_id="AAG55997.1"  
/db\_xref="GI:12514834"

gene  
CDS  
1443..1787  
/gene="21906"  
/function="orf; Other or unknown (phage or prophage related)"  
/note="Residues 1 to 113 of 114 are 49.12 pct identical to 916901599|gb|AF31102.1| (AF069529) gp11 [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"  
/protein\_id="AAG55998.1"  
/db\_xref="GI:12514835"

gene  
CDS  
1784..2572  
/gene="21908"  
/function="putative structure; Structural component (phage or prophage related)"  
/note="Residues 25 to 258 of 262 are 67.09 pct identical to 916901588|gb|AF31091.1| (AF069529\_4 (AF069529) major tail subunit [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="putative tail component of prophage CP-933X"  
/protein\_id="AAG56002.1"  
/db\_xref="GI:12514839"

CDS

gene  
CDS  
2578..2820  
/gene="21910"  
/function="orf; Other or unknown (phage or prophage related)"  
/note="Residues 1 to 77 of 80 are 62.33 pct identical to 916901601|gb|AF31104.1| (AF069529) gp14 [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"  
/protein\_id="AAG56000.1"  
/db\_xref="GI:12514837"

gene  
CDS  
2978..3259  
/gene="21912"  
/function="orf; Other or unknown (phage or prophage related)"  
/note="Residues 1 to 92 of 93 are 60.86 pct identical to 916901601|gb|AF31104.1| (AF069529) gp14 [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"  
/protein\_id="AAG56001.1"  
/db\_xref="GI:12514838"

gene  
CDS  
3307..6387  
/gene="21913"  
/function="putative structure; Structural component (phage or prophage related)"  
/note="Residues 1 to 1022 of 1026 are 52.47 pct identical to 916901589|gb|AF31092.1| (AF069529\_5 (AF069529) tail length tape measure protein [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="putative tail component of prophage CP-933X"  
/protein\_id="AAG56002.1"  
/db\_xref="GI:12514839"

gene  
CDS  
3307..6387  
/gene="21913"  
/function="putative structure; Structural component (phage or prophage related)"  
/note="Residues 1 to 1022 of 1026 are 52.47 pct identical to 916901589|gb|AF31092.1| (AF069529\_5 (AF069529) tail length tape measure protein [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=11  
/product="putative tail component of prophage CP-933X"  
/protein\_id="AAG56002.1"  
/db\_xref="GI:12514839"

Query Match 4.6%: Score 22: DB 1: Length 7286;  
 Best Local Similarity 100.0%: Pred. No. 0.47;  
 Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 TAATCGGACACCGGTAAACAC 312  
 ||||||||||||||||||  
 Db 1242 TAATCGGACACCGGTAAACAC 1263

RESULT 7  
 AE005325  
 LOCUS  
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 149  
 OF 155.  
 ACCESSION AE005325  
 VERSION AE005325.1  
 KEYWORDS GI:12514723  
 SOURCE Escherichia coli O157:H7 EDL933.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE  
 AUTHORS 1 (bases 1 to 10029)  
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,  
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,  
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,  
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,  
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
 Welch,R.A. and Blattner,F.R.  
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7  
 Nature 409 (6819), 529-533 (2001)

TITLE  
 JOURNAL MEDLINE  
 PUBMED 11206551

REFERENCE  
 AUTHORS 2 (bases 1 to 10029)  
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,  
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,  
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,  
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,  
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
 Welch,R.A. and Blattner,F.R.  
 Direct Submission  
 Submitted (22-OCT-2000) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

TITLE  
 JOURNAL  
 FEATURES  
 source  
 misc\_feature  
 gene  
 CDS  
 gene  
 CDS

/gene="21809"  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 /note="Residues 1 to 148 of 148 are 75.16 pct identical to  
 residues 1 to 149 of 149 from GenPept 118 :  
 g116901598|gb|AAE31101.1| (AF069529) Gp10 [Bacteriophage  
 HK97]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933N"  
 /protein\_id="AAG55909.1"  
 /db\_xref="GI:12514725"  
 /translation="MTETLDFSGLEDISDQLDLSGAENRRLREATRAGANVLKEE  
 VSRARVRKRLRRNVVLSRRSRDGMESGVIRVNPDTGSDNTMKADNRFNAFY  
 WRFEVGVNMPHPVPRFPAFVDVRSFOAAOVLAIRKMRALDEVLRR"  
 752..1096  
 /gene="21810"  
 752..1096  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 /note="Residues 1 to 113 of 114 are 49.12 pct identical to  
 residues 1 to 114 of 115 from GenPept 118 :  
 g116901599|gb|AAE31102.1| (AF069529) Gp11 [Bacteriophage  
 HK97]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933N"  
 /protein\_id="AAG55910.1"  
 /db\_xref="GI:12514726"  
 /translation="MTPEADLYPLAHLAGGOVYPVYVPLIDGRPSVALPWNVPSLSS  
 VSDVWGGQAESSVQIDVYACTVQARQIRQDAREALMLAPGSVSEMDYIPENR  
 CYRATLEFOYTV"  
 1093..1878  
 /gene="21811"  
 1093..1878  
 /function="putative structural component (Phage  
 or Prophage Related)"  
 /note="Residues 24 to 257 of 261 are 67.09 pct identical  
 to residues 1 to 234 of 234 from GenPept 118 :  
 g116901588|gb|AAE31091.1| (AF069529)\_4 (AF069529) major tail  
 subunit [Bacteriophage HK97]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail component encoded by prophage  
 CP-933N"  
 /protein\_id="AAG55911.1"  
 /db\_xref="GI:12514727"  
 /translation="MTFSFYKTHTPRAGFLISGGRMSALYRSOLTOVMSSAPRT  
 AETMDKAEYLRDCTTKEVOFTAGOKODIDVTTCSTEDENINGLGASSISMSGNFY  
 LNOAQNLRDAYNDALYARVLEPSSGKRFIAEVRQHTWSSGTNGVVAATSLRK  
 GKPSFVPLAPFKNPDKLTVTNGALLTWSVSVNGSTPPYKIHAKMKKDGOPVEGQTTD  
 TFSKPGQASADACKYTCVTVDSAEKQSVTSVACTVTVSAAG"  
 1845..2267  
 /gene="21812"  
 1845..2267  
 /function="putative factor; Structural component (Phage or  
 prophage Related)"  
 /note="Residues 17 to 140 of 140 are 62.90 pct identical  
 to residues 1 to 123 of 123 from GenPept 118 :  
 g116901601|gb|AAE31104.1| (AF069529) Gp14 [Bacteriophage  
 HK97]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail assembly chaperone encoded by  
 prophage CP-933N"  
 /protein\_id="AAG55912.1"  
 /db\_xref="GI:12514728"  
 /translation="MHRDSERSRRIRGWIIMKDKLTALRLSGFRHKTKVYPENRN  
 VSVLRPSAEAWYLMQEVYNGDGDGDDTLTSVAKTRRNLEADVTLFCDVLCOTDQOR  
 VTFPDDBQVLAIVGPNHARLLRQALELTADESARKK"

```

gene
2291..2572
/gene="21813"
CDS
2291..2572
/gene="21813"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 90 of 93 are 61.11 pct identical to
residues 133 to 224 from Genpept 118 :
g116901591gb|AAE31104.11 (AF069529) Gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933N"
/protein_id="AAG55913.1"
/db_xref="GI:12514729"
/translation="MRLRLGRLSELRLSLSASAMMMEDFRVSLGDERGDIIN
AQIVAVGAGQGMVNLKDALMCWGEDDEKPEVDPALEDAISLAAMS"
2624..3817
/gene="21814"
2624..3817
/gene="21814"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 393 of 397 are 69.36 pct identical to
residues 1 to 394 of 1089 from Genpept 118 :
g116901589gb|AAE31092.11AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933N"
/protein_id="AAG55914.1"
/db_xref="GI:12514730"
/translation="MATLRLIRIKSANSQSFSEIORSRMGSEYRTLONGROAA
AAARORRALNELNSDLTEIRGSVAGAFAGAFSGHLISLAEMSSVAKROAS
KSGASTAENGSVITQFSGALAGVLRGEPNSVNESDRVLAAGMVARKDK
AMADGKRLADKVPALISQGLRDEYAAKEVSSITTEVNAFMVAGANEASG
VTKLSGMLNGVAGOIDNATGAVATGAVATGAVATGAVATGAVATGAVATG
LAEOFRGTQIATARAARAAYRAQQAARQETEMQIAEARLAATGERLNRIANRT
AAPG"
4038..4736
/gene="21815"
4038..4736
/gene="21815"
/function="putative structure; structural component (Phage
or Prophage Related)"
/note="Residues 1 to 232 of 232 are 76.72 pct identical to
residues 1 to 232 of 232 from Genpept 118 :
g112151221gb|AAE6550.11 (U02459) L (tail component;232)
[bacteriophage lambda]"
/codon_start=1
/transl_table=11
/product="putative tail component L homolog encoded by
prophage CP-933N"
/protein_id="AAG55915.1"
/db_xref="GI:12514731"
/translation="MONIHESLINESVSESGSPRIVLWIDLVQGGRRYFECNELNE
KGEAVTWGQROYVYPLDGSPEMNGKSSARSLTVNSLGLVTGMAEDIOSVGT
VVRARVARELDVAVNAGNPEADPDELRODMVVEOMSELTATATASFAVATPEIDG
ALPGRIMLANMDMYRDECGYHGPVAVDEFPNPTDIRKDRCSKCMGCEMRGAV
ANFGGSLINKLSQ"
4911..5069
/gene="21816"
4911..5069
/gene="21816"
/function="orf; Other or unknown (Phage or Prophage
Related)"
Query Match 4.6%; Score 22; DB 1; Length 10029;
Best Local Similarity 100.0%; Pred No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 291 TAATCCGACACCGGTAAACACG 312
|||||

```

```

DB 551 TAATCCGACACCGGTAAACACG 572
RESULT 8
AE005290
LOCUS
DEFINITION
AE005290 11809 bp DNA linear BCT 21-MAR-2001
Escherichia coli O157:H7 EDL933 genome, config 1 of 3, section 114
of 155.
ACCESSION
AE005290 AE005174
VERSION
AE005290.1 GI:12514210
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7 EDL933.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 11809)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
11206551
REFERENCE
2 (bases 1 to 11809)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
11809
Location/Qualifiers
1..11809
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
<1..>11809
/note="cryptic prophage CP-933M; includes one copy of the
60 bp direct repeat that flanks the prophage"
<1..>11809
/note="O-island #44; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
109..312
/gene="21367"
109..312
/gene="21367"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 67 of 67 are 56.71 pct identical to
residues 48 to 112 of 112 from Genpept 118 :
g116901597gb|AAE31100.11 (AF069529) putative head-tail
adaptor [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG55500.1"
/db_xref="GI:12514211"
/translation="MQSGAEMQAEIRIWRGSGREITRAASRLHYLSGPMWRDILNV
VGLPVDATVGGRLRLICRLGEGK"
309..755
/gene="21368"
309..755
/gene="21368"
/function="orf; Other or unknown (Phage or Prophage
Related)"

```

```

/note="Residues 1 to 148 of 148 are 75.16 pct identical to
residues 1 to 149 of 149 from GenPept 118 :
g116901598|gb|AAJ31101.1| (AF069529) Gp10 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5501.1"
/db_xref="GI:12514212"
/translation="MTELLDESGLEDISRDQLQLSGAENRNYLRATAGANVLEE
VVSRAPIRGRKMLRRNVVLSRRSGDMESGVIIRGNPDGTMSDNTMRADNRNMFY
WRFVEMGTVMRPHFVRPAFVRSQAQVAIAIRNRAIDEVLR"
/size..1096
/gene="21369"
/size..1096
/gene="21369"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 113 of 114 are 48.24 pct identical to
residues 1 to 114 of 115 from GenPept 118 :
g116901599|gb|AAJ31102.1| (AF069529) Gp11 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5502.1"
/db_xref="GI:12514213"
/translation="MTEADLYPHLAHLAGGVYVYVPLDGRPSVALPMVVFSLSS
VSADVMGGAESSVQIBVYAGTVQARQIRQDAREXIMLLAPGSVSEMDVPEPNNR
CYRATLEPROTV"
/size..1872
/gene="21370"
/size..1872
/gene="21370"
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 22 to 255 of 259 are 66.66 pct identical
to residues 1 to 234 of 234 from GenPept 118 :
g116901588|gb|AAJ31091.1|AF069529_4 (AF069529) major tail
subunit [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="putative tail component encoded by cryptic
prophage CP-933M"
/protein_id="AAG5503.1"
/db_xref="GI:12514214"
/translation="MNFPRTHNPPRAGFLLSGRMSALYERSOLQVMISSAPATAE
TMKAYLRIDCTIKVOPFAGQKODIDYTLCTROENINGIASSETMSGNFYLN
QONALRLDAYINDNTVAFKVPQPSGKGFPLAEVRQHTSSGNGVAAFTSLRLKLG
PVSIVPLAFVKNLDTLTVTGALLTMSVSVNGGTPPYKHAHKDKGQPEGGTDTJF
SKANTSGDKGAVTCEVTDABQPOSITSDACTVTNAGAG"
/size..2252
/gene="21371"
/size..2252
/gene="21371"
/size..2252
/gene="21371"
/function="putative factor; Structural component (Phage or
Prophage Related)"
/note="Residues 1 to 124 of 124 are 62.90 pct identical to
residues 1 to 123 of 124 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) Gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="putative tail assembly chaperone encoded by
cryptic prophage CP-933M"
/protein_id="AAG5504.1"
/db_xref="GI:12514215"
/translation="MAKDLKTLALARLSEFRHKTVKVPPEMRNVSVLREPSAEAMYLW
QEVNLNGDGBDDDTLSVAKTRNLLEADVTLCFCDVLCDTDLQRFVAFDDREOVLAVYGP
VHARLLRQALELTLADAESARKK"
/size..2276
/size..2276

```

```

CDS
/gene="21372"
/size..2276
/size..2276
/gene="21372"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 90 of 93 are 60.00 pct identical to
residues 133 to 222 of 224 from GenPept 118 :
g116901601|gb|AAJ31104.1| (AF069529) Gp14 [Bacteriophage
HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5505.1"
/db_xref="GI:12514216"
/translation="MRLALRLGRTLSLRLHSLSASAFAMMMERFDRVSPDGEDIRN
AQIVKAVFGAQCMMVALKDALMLCWMGEDEKPEVDPAALDALSLAMS"
/size..3610
/gene="21373"
/size..3610
/gene="21373"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 241 of 260 are 79.66 pct identical to
residues 71 to 311 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5506.1"
/db_xref="GI:12514217"
/translation="MXGAFXGAFATGHLISLADWSSVNXRLKQAOSSDEFSSXOKV
LMDISQRTGXAESDNALFXRSPASMRREYGSADVLKTEAISTGLKISGASTABAG
SVTQPSQALAQGLRGEFNSVNESGDRIVRLAAGMVARDLKAMADDGOLFADK
VVALISOLEVLEDEYEAAMPETVSDGITREYENAFMAMVGANFASVTVTLISGLNGV
ACQIDNVATARGTGCRRCRCPVLMQYGLASDVCGRCTCGCT"
/size..3477
/gene="21374"
/size..3477
/gene="21374"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 14 to 107 of 111 are 35.41 pct identical
to residues 300 to 394 of 1089 from GenPept 118 :
g116901589|gb|AAJ31092.1|AF069529_5 (AF069529) tail length
tape measure protein [Bacteriophage HK97]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933M"
/protein_id="AAG5507.1"
/db_xref="GI:12514218"
/translation="MLPDRLLIMQOPVGLAVGVARFYFGNMGASMAPAGLVTAAR
NEVALAEQFRGTQITARARARXAVYRAQQAVALAARKEEMQIAEARLALTQERLNNI
AKRTAAPQC"
/size..4033
/size..4033
/gene="21375"

```

Query Match 4 6%: Score 22; DB 1; Length 11809;  
 Best Local Similarity 100.0%; Pred. No. 0.48;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCCGACACCGGTAACAGC 312  
 |||||  
 Db 551 TAATCCGACACCGGTAACAGC 572

RESULT 9  
 AE005441/c  
 LOCUS  
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 60

290. 05441.1 AE005174  
05441.1 GI:12516368

hierarchy coli 0157:H7 EDL933.  
terial: Proteobacteria; gamma subdivision; Enterobacteriaceae; *Escherichia*.

bases 1 to 13505

1. 13505  
Location/Qualifiers  
organism="Escherichia coli 0157:H7 EDL933"  
strain="EDL933"  
serotype="O157:H7"  
db\_xref="taxon:155864"  
note="enterohemorrhagic"  
1. >13505  
/note="O-island #93: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933V: includes one copy of the 21 bp direct repeat that flanks the prophage; encodes stx1a and stx1b subunits of a Shiga-like toxin"  
complement(198..374)  
/gene="23316"  
complement(198..374)  
/gene="23316"  
function="orf: Other or unknown (phage or Prophage Related)"  
note="No significant matches"  
/codon\_start=1  
/product="unknown protein encoded within prophage CP-933v"  
/protein\_id="AA657205.1"  
/db\_xref="GI:12516369"  
translation="MNGSAGLVSPVYHAIQNDGNKGQIGPAAAGSLVQLDQHQVA VMISRRADSGMISG"  
complement(371..3280)  
/gene="23318"  
complement(371..3280)  
/gene="23318"  
/function="putative structure: Structural component (phage or Prophage Related)"  
note="Residues 1 to 962 of 969 are 53.92 pct identical to residues 1 to 1015 of 1089 from GenPept 118 : g116901581|gb|AAE31092.1|AF069529.5 (AF069529) tail length type measure protein [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=1  
/product="putative tail component of prophage CP-933v"  
/protein\_id="AA657206.1"  
/db\_xref="GI:12516370"  
translation="MATELLIKTISANSQSFSEIORASMGSEYRTTLQNGSOAA AANEGRALAEINLSQTEIRNSAVGAGAPAGAFASGHILISLADWSSVNRKIQAS

290. 05441.1 AE005174  
05441.1 GI:12516368

hierarchy coli 0157:H7 EDL933.  
terial: Proteobacteria; gamma subdivision; Enterobacteriaceae; *Escherichia*.

bases 1 to 13505

1. 13505  
Location/Qualifiers  
organism="Escherichia coli 0157:H7 EDL933"  
strain="EDL933"  
serotype="O157:H7"  
db\_xref="taxon:155864"  
note="enterohemorrhagic"  
1. >13505  
/note="O-island #93: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933V: includes one copy of the 21 bp direct repeat that flanks the prophage; encodes stx1a and stx1b subunits of a Shiga-like toxin"  
complement(198..374)  
/gene="23316"  
complement(198..374)  
/gene="23316"  
function="orf: Other or unknown (phage or Prophage Related)"  
note="No significant matches"  
/codon\_start=1  
/product="unknown protein encoded within prophage CP-933v"  
/protein\_id="AA657205.1"  
/db\_xref="GI:12516369"  
translation="MNGSAGLVSPVYHAIQNDGNKGQIGPAAAGSLVQLDQHQVA VMISRRADSGMISG"  
complement(371..3280)  
/gene="23318"  
complement(371..3280)  
/gene="23318"  
/function="putative structure: Structural component (phage or Prophage Related)"  
note="Residues 1 to 962 of 969 are 53.92 pct identical to residues 1 to 1015 of 1089 from GenPept 118 : g116901581|gb|AAE31092.1|AF069529.5 (AF069529) tail length type measure protein [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=1  
/product="putative tail component of prophage CP-933v"  
/protein\_id="AA657206.1"  
/db\_xref="GI:12516370"  
translation="MATELLIKTISANSQSFSEIORASMGSEYRTTLQNGSOAA AANEGRALAEINLSQTEIRNSAVGAGAPAGAFASGHILISLADWSSVNRKIQAS

290. 05441.1 AE005174  
05441.1 GI:12516368

hierarchy coli 0157:H7 EDL933.  
terial: Proteobacteria; gamma subdivision; Enterobacteriaceae; *Escherichia*.

bases 1 to 13505

1. 13505  
Location/Qualifiers  
organism="Escherichia coli 0157:H7 EDL933"  
strain="EDL933"  
serotype="O157:H7"  
db\_xref="taxon:155864"  
note="enterohemorrhagic"  
1. >13505  
/note="O-island #93: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933V: includes one copy of the 21 bp direct repeat that flanks the prophage; encodes stx1a and stx1b subunits of a Shiga-like toxin"  
complement(198..374)  
/gene="23316"  
complement(198..374)  
/gene="23316"  
function="orf: Other or unknown (phage or Prophage Related)"  
note="No significant matches"  
/codon\_start=1  
/product="unknown protein encoded within prophage CP-933v"  
/protein\_id="AA657205.1"  
/db\_xref="GI:12516369"  
translation="MNGSAGLVSPVYHAIQNDGNKGQIGPAAAGSLVQLDQHQVA VMISRRADSGMISG"  
complement(371..3280)  
/gene="23318"  
complement(371..3280)  
/gene="23318"  
/function="putative structure: Structural component (phage or Prophage Related)"  
note="Residues 1 to 962 of 969 are 53.92 pct identical to residues 1 to 1015 of 1089 from GenPept 118 : g116901581|gb|AAE31092.1|AF069529.5 (AF069529) tail length type measure protein [Bacteriophage HK97]"  
/codon\_start=1  
/transl\_table=1  
/product="putative tail component of prophage CP-933v"  
/protein\_id="AA657206.1"  
/db\_xref="GI:12516370"  
translation="MATELLIKTISANSQSFSEIORASMGSEYRTTLQNGSOAA AANEGRALAEINLSQTEIRNSAVGAGAPAGAFASGHILISLADWSSVNRKIQAS

CDS	/gene="23323" complement(4793..5137) /gene="23323" /function="orf: Other or unknown (phage or Prophage Related)" /note="Residues 1 to 113 of 114 are 49.12 pct identical to residues 1 to 114 of 115 from Genpept 118 : g1169015981gb AAF31102.11 (AF069529) Gp11 [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="unknown protein encoded within prophage CP-933v" /protein_id="AAG57210.1" /db_xref="GI:12516374" /translation="MTEADLYPHLAHLAAGGYPPVYVPLDGRPSVALPWPVFLISSVSADVWGGAESSVSVDIVAGTVTAQRQIKQDAREALIMLLAPGSVSEMODYIPENRCYATLEFQVY" complement(5134..5580) /gene="23325" complement(5134..5580) /gene="23325" /function="orf: Other or unknown (phage or Prophage Related)" /note="Residues 1 to 148 of 148 are 75.16 pct identical to residues 1 to 149 of 149 from Genpept 118 : g1169015981gb AAF31101.11 (AF069529) Gp10 [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="unknown protein encoded within prophage CP-933v" /protein_id="AAG57211.1" /db_xref="GI:12516375" /translation="MTEFLDPSGLEDISRDQLLSGAEENNRVLEATRAGANVLEEVSRAVRRGLKRRNVVLSRRSGDGMESGVHICGVNPDTGNSDVTMTKADNPRNAFYWRFVEMGTVMPPHPVPAPFVRSQQAAQVAIAIRNRAIDVLR" complement(5577..5927) /gene="23326" complement(5577..5927) /gene="23326" /function="orf: Other or unknown (phage or Prophage Related)" /note="Residues 6 to 116 of 116 are 45.94 pct identical to residues 4 to 112 of 112 from Genpept 118 : g1169015971gb AAF31100.11 (AF069529) putative head-tail adaptor [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="unknown protein encoded within prophage CP-933v" /protein_id="AAG57212.1" /db_xref="GI:12516376" /translation="MAISAGRLTOMISVLPVLTNRNAGEMTEEMVSCGIHADIRGRSSRRMQSGAEQAQAEIRIWRGQSGRETTAASRLCHLVLSGPMRDXXLVNVGXPVPDXT	4.6%; Score 22; DB 1; Length 13505; Best Local Similarity 100.0%; Pred. No. 0.48; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Query Match 291 TAAATCGGACACCGGTACAGC 312       Db 5338 TAAATCGGACACCGGTACAGC 5317	RESULT 10 AE006459/c 13624 bp DNA 1linear BCT 21-MAR-2001 LOCUS Escherichia coli O157:H7 EDL933 genome, contig 2 of 3, section 53 DEFINITION of 55. ACCESSION AE006459 AE005174 VERSION AE006459.1 GI:13259579 KEYWORDS SOURCE Escherichia coli O157:H7 EDL933. ORGANISM Escherichia coli O157:H7 EDL933 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
gene	/gene="26035" complement(3488..3769) /gene="26035" /function="putative factor: DNA packaging, phage assembly (Phage or Prophage Related)" /note="Residues 1 to 92 of 93 are 60.86 pct identical to residues 133 to 224 of 224 from Genpept121: gb AAF31104.11 (AF069529) Gp14 [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="putative tail assembly protein of cryptic prophage CP-933p" /protein_id="AAK16951.1"	gene CDS	REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS JOURNAL FEATURES source Escherichia. 1 (bases 1 to 13624) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoustis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R. Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 Nature 409 (6819), 529-533 (2001) 2 (bases 1 to 13624) Plunkett,G. III. Direct Submission Submitted (12-FEB-2001) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706-1577, USA Location/Qualifiers 1. 13624 /organism="Escherichia coli O157:H7 EDL933" /strain="EDL933" /serotype="O157:H7" /db_xref="taxon:155864" /note="enterohemorrhagic" complement(198..3440) /gene="26034" complement(198..3440) /gene="26034" /function="putative structure: Structural component (phage or Prophage Related)" /note="Residues 1 to 1076 of 1080 are 52.37 pct identical to residues 1 to 1079 of 1080 from Genpept121: gb AAF31092.1 AF069529_5 (AF069529) tail length tape measure protein [Bacteriophage HK97]" /codon_start=1 /transl_table=11 /product="putative tail component of cryptic prophage CP-933p" /protein_id="AAK16950.1" /db_xref="GI:13259580" /translation="MATLERLIKITSANSQSFQSEIQRASRMGSEYRTLQNGRQAA AAAREQRALAEIHSQTEIRASAVMTGAFGATGHLISLADMSVNRRLQAS OSSDERASSOKVIMDISORTGTAESDNALFPRSASMEYGVSDLYKVEALSTG LITSGASTAEAGSVITPQFSGALACQVLRDEEFTNSVNSGDRIVRAALAAQMGARADLK AMADDKLTADKYVPALISQGLILREYRAMEYTSSTITKENAFMAVGCANESAG VKTLSGMLNGVAGQIDNVATVAGVAVARYRGNMMSGAMSAATAGLVTAARNEVA LAEQFRGTQIATARAARAVYRAQQAVALAARQTEMOIAEARLAATQERLNRIARS AAONALNSTTAVGSRMLSGALGVGPGVLMGLAAAVTLKXONBOAREASROYALT IDEIAHKTPMSLSPKASDNGEGRRAALTRPONHLIDQASRVSLOEKINGYQYVLANP GMTTGGGFEINHLTSKYTYTBELAQTEDQLAYEQSRLQMOEKAOSIDQVLAGLEDR VALIRQOAAQKRVYOSMLVMGQTYTEENRLLGLGELLQQRGLVNVFLRPLQATLD DKQOQSLITTERELALSLRKGKEKERVRLGYAADLGFVGDPEYQERORYISNALAEAW RNNENVKPKSRGKSETERKEDSFSLRLOQKEQLALACQNTLEAKLYOTALGELKT LSEIQOELLRNALALIDQOKIREQLRYREETLKNVVARASNESELGYGGEARARE RMRELOQIRDSFRKQADLOSOYOTGDISEDPYRQALQANAOYLISRLKIDQVFAERS DYQRADMQKIDEGFSNWNVDNSASQAQALATBGISMWNITFMLEKNGVAFRSM ASSVLOEISKVLNNAIVNGITTAAGSKAGAGFUSITGDMUGAANAKGCVITTSAN LSAYSNSIVDTPYEFARFAGAGLAGEAGEAEMPIITRAADGSLGVAVASNMGSAGLV VSPVYHAIQNDGANGQIGPEAAGSLVQIDQVAVAVLMSRRDGMLSG"	

```

/ob_xref="GI:13259581"
/translation="MRLALRIGRTLSLSLSASEAMMMMEFRISPLGDDRGDIRN
AQIVKAVGAGGMNVALKDAMLGWGEDEDEKPEVDPPALEDAISLAAGS"
complement(3793..4215)
/gene="26036"
complement(3793..4215)
/gene="26036"
/function="putative factor: DNA packaging, phage assembly
(PHAGE OR PHAGE RELATED)"
/note="Residues 17 to 140 of 140 are 62.90 pct identical
to residues 1 to 123 of 224 from Genpept121:
gb|AA069529.1| (AF069529) gp14 [Bacteriophage HK97]"
/codon_start=1
/translation=1
/product="putative tail assembly protein of cryptic
prophage CP-933p"
/protein_id="AAK16952.1"
/ob_xref="GI:13259582"
/translation="MHRDSESRIRGIVNKKDKTLALARLSGFRHKTVKPEGN
VSVLRPESEAWYIMOVINGDEDDDTLSVAKTRRNLEADVTLFCVYLCDDLDLR
VFPPDREOVAVYGPVHARLLKOLELIADAESARKK"
complement(4182..4967)
/gene="26037"
complement(4182..4967)
/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 24 to 257 of 261 are 68.37 pct identical
to residues 1 to 234 of 234 from Genpept121:
gb|AA069529.1| (AF069529) major tail subunit
[Bacteriophage HK97]"
/codon_start=1
/translation=1
/product="putative tail component of cryptic prophage
CP-933p"
/protein_id="AAK16953.1"
/ob_xref="GI:13259583"
/translation="MPSFPRYKTHPPRAGFLSGRMSALYERISOLTOVMISATP
AETMDKAEYLRLDCTIKETKQYFAGOKODIVTTLCTSEONINGASSISNGFV
LNOAGALRDADNDALAFVLEPSSGGEFLAEVROHMTSSGNGVAAATPSLTK
GKPSVNVPLATVKNLDTLTNTGALLTMSVANGCTPPYKAMKRDQDPVDCQDTL
TESKFGAQSADGKTYCTVNTSAEKAQSVTSVECTYVSAAG"
complement(4964..5308)
/gene="26038"
complement(4964..5308)
/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 1 to 113 of 114 are 49.12 pct identical to
(A069529) gp11 [Bacteriophage HK97]"
/codon_start=1
/translation=1
/product="putative structural component of cryptic
prophage CP-933p"
/protein_id="AAK16954.1"
/ob_xref="GI:13259584"
/translation="MTEADLYPHLAHLAGGVYVYVLLDGRSVALLPWPVSLISS
VSADVMGGAESSVSVDVYAGTVTAQRIROAREINMLLAPGVSEMODYIPENR
CYRATIEFOVY"
complement(5305..5751)
/gene="26039"
complement(5305..5751)
/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 1 to 148 of 148 are 75.16 pct identical to
(A069529) gp10 [Bacteriophage HK97]"
/codon_start=1
/translation=1
/product="putative structural component of cryptic
prophage CP-933p"

```

```

/gene="26040"
complement(5748..6005)
/gene="26040"
/function="putative structure: Structural component (Phage
or Prophage Related)"
/note="Residues 6 to 85 of 85 are 53.75 pct identical to
(A069529) putative head-tail adaptor [Bacteriophage
HK97]"
/codon_start=1
/translation=1
/product="putative head-tail adaptor of cryptic prophage
CP-933p"
/protein_id="AAK16956.1"
/ob_xref="GI:13259586"
/translation="MSCGKIHADIRGSSRRRMOGSAEMAQAEIRIWRGSGRETTA
ASRLHVLSPWMDRLTNVGLPVPDATTGRLTLCRLGCKR"
complement(6108..6434)
/gene="26041"
complement(6108..6434)
/function="orf: Other or unknown (Phage or Prophage
Related)"
/note="Residues 5 to 25 of 108 are 66.66 pct identical to
residues 6 to 26 of 104 from Genpept121:
gb|AA069529.1| (AF198256_14) (AF198256) unknown [Haemophilus
influenzae]"
/codon_start=1
/translation=1
/product="unknown protein encoded by cryptic prophage
CP-933p"
/protein_id="AAK16957.1"
/ob_xref="GI:13259587"
/translation="MTALLTLEELRAHLRVDHADDMDLNDKVRQATAVLLAYIOGSR
DKVIRDDGELIPGEALTRMKAGAMRLTGMILYRNPDLAERBELIQGELPSSVSLIYDL
RCPTVY"
complement(6437..9016)
/gene="26042"
complement(6437..9016)

```

```

Query Match
Best Local Similarity 4.6% Score 22: DB 1: Length 13624;
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 291 TAATCCGACACCGGTAAACAGC 312
DB 5509 TAATCCGACACCGGTAAACAGC 5488

```

```

RESULT 11
AL139161
LOCUS
DEFINITION
Human DNA sequence from clone Rf5-940F7 on chromosome 1q42.2-43,
complete sequence.
ACCESSION
AL139161
VERSION
AL139161.8 GI:15131780
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 103608)
AUTHORS
Martin, S.
TITLE
Direct Submission
JOURNAL
Submitted (04-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk; Clone
requests: clonerequest@sanger.ac.uk

```



## COMMENT

On Aug 9, 2001 this sequence version replaced g1:14787245. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further Information can be found at <http://www.sanger.ac.uk/HGP/Chrl> RP5-940F7 is from the library RPCI-5 constructed by the group of Pleter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-940F7. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP4-764D2 is at 101609 in this sequence. The true right end of clone RP4-580N22 is at 2000 in this sequence.

## FEATURES

## source

1.103608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q42.2-43"  
/clone="RP5-940F7"  
/clone\_lib="RPCI-5"  
1.151  
/note="AluSg/X repeat: matches 145. .295 of consensus"  
369. .935  
/note="L1PB2 repeat: matches 5574. .6155 of consensus"  
936. .1247  
/note="AluSp repeat: matches 1. .313 of consensus"  
1248. .3222  
/note="L1PB2 repeat: matches 3614. .5574 of consensus"  
2001. .2044  
/note="Single clone region. Assembly confirmed by restriction digest data."  
3462. .3798  
/note="MER58B repeat: matches 2. .341 of consensus"  
4257. .4555  
/note="AluSc repeat: matches 5. .298 of consensus"  
4574. .4663  
/note="AluSc repeat: matches 161. .248 of consensus"  
4714. .4753  
/note="L2 repeat: matches 161. .248 of consensus"  
4812. .4870  
/note="L2 repeat: matches 161. .248 of consensus"  
4870. .5107  
/note="MER4D repeat: matches 359. .417 of consensus"  
5895. .6160  
/note="MER4D repeat: matches 9. .234 of consensus"  
5895. .6160  
/note="AluSx repeat: matches 6. .285 of consensus"  
6161. .6196  
/note="L2 repeat: matches 6. .285 of consensus"  
6290. .6402  
/note="L2 repeat: matches 6. .285 of consensus"  
6290. .6402  
/note="L2 repeat: matches 6. .285 of consensus"  
6837. .6868  
/note="L2 repeat: matches 6. .285 of consensus"  
7155. .7456  
/note="L2 repeat: matches 6. .285 of consensus"

repeat\_region 7653. .7833  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 8397. .8646  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 8954. .9255  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 9373. .9506  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 10071. .10761  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 10847. .11024  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 11025. .11251  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 11275. .11414  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 12228. .12536  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 12537. .12829  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 13542. .13637  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 14166. .14457  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 15312. .15622  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 16455. .16515  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 17415. .17458  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 17480. .17949  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 17950. .18049  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 18444. .18636  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 18759. .18895  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 19595. .19610  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 19611. .19921  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 19922. .19993  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 20033. .20134  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 21253. .21557  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 21740. .21793  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 21799. .22106  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 22112. .22408  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 23607. .23906  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 24101. .24403  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 25145. .25560  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 25713. .26011  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 26525. .26570  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 26573. .26859  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 27151. .28279  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 28500. .28621  
/note="L2 repeat: matches 33. .208 of consensus"  
repeat\_region 28659. .28872  
/note="L2 repeat: matches 33. .208 of consensus"

```

repeat_region /note="AluIo repeat: matches 85. .287 of consensus"
29508. .29715 /note="MER30 repeat: matches 1. .230 of consensus"
repeat_region /note="MER30 repeat: matches 1. .230 of consensus"
29785. .30063 /note="AluIo repeat: matches 1. .250 of consensus"
repeat_region /note="LIM4 repeat: matches 4494. .4608 of consensus"
30640. .30754 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
30762. .30933 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
30942. .31167 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
31192. .32894 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
31194. .33485 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
33563. .33697 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
33698. .33995 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
33996. .34174 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
34670. .34769 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
34927. .35112 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
35604. .35905 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
35906. .36181 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
36585. .36625 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
36626. .36797 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
36795. .38360 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
38361. .38657 /note="LIM4 repeat: matches 1. .182 of consensus"
repeat_region /note="LIM4 repeat: matches 1. .182 of consensus"
38658. .41845 /note="LIM4 repeat: matches 1. .182 of consensus"

```

```

Query Match
Best Local Similarity 100.0%; Score 22; DB 9; Length 103608;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 257 AAAAAAGTAAAGCGCGCAGCT 278
Db 7141 AAAAAAGTAAAGCGCGCAGCT 7162

```

```

RESULT 12
AP002555 222605 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 6/20.
ACCESSION AP002555 BA000007
VERSION AP002555.1 GI:13360886
KEYWORDS
SOURCE

```

```

ORGANISM
DNA
Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 050952)
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Sasakawa,C., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Complete nucleotide sequence of the prophage VP2-Sakai carrying the
derived from the sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and

```

```

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

```

```

TITLE
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
JOURNAL
MEDLINE
REFERENCE
AUTHORS
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Shinagawa,H., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Complete nucleotide sequence of the prophage VP2-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsuda,E., Nakayama,K., Murata,T.,
Tanaka,M., Toke,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 222605)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
Location/Qualifiers
1. .222605
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 050952"
/db_xref="taxon:83334"
74. .607
/gene="ECS1423"
74. .607
/gene="ECS1423"
/gene="ECS1423"
/note="Similar to B1045.ECOLI g11787283 percent identity
100 in 177 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence-not-experimental
/product="putative polyprotein"
/db_xref="GI:13360887"
/translation="MKTIRHVVQGDITKLAVDYVYNNANSLMGCGVDCGAIHRAAP
NSLRVIAANSYVAFPAISGVGYPRAAAEIAVATVSEFTIRHALPEQVYFCYD
EENVAHYERLILQOCDE"
549. .2030
/gene="ECS1424"
549. .2030
/gene="ECS1424"
/note="Similar to YMDC.ECOLI g11787284 percent identity
99 in 493 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence-not-experimental
/product="putative synthase"
/protein_id="BA34487.1"
/db_xref="GI:13360888"
/translation="MKTPTSTKDSLPKKNLPLASAVPLCSHPGCGGLFPL
EKSIDAPAAKYRKAEMSEHTLDVQYIWMDDMSGILFSAIAAKAGVAVRLLDN
NTPGCDLIDRLDSHPRIEVRLEPNFSLRLPLGITYITDPSRLNRHMKSFVQGVV

```

```

COMMENT
FEATURES
SOURCE
gene
CDS

```

TGVGADNADIEYFAGNGEPELESDIDLMAGSPVEVDADPEARWYCKSVSPLIDSDVDPK  
 PLGEKRNADITELPAGNDNDAMTRITLRMAKSSPEPTINADCTPEITNAKTRITLSDVDPK  
 GEGKAKRSLLEPQRLIEDNGSPSERIDIISSYFVPVPRAGVAQOLLNWRKVKIATILN  
 SLADNADVAVNGAWRKMRKLLRYNGVELTELPEPTROSSYTLHBRGITNGASGIAHAT  
 FSIDKQVYVIGENFDRPSTLNTENGFPVIESPTLQDLIDKRFIOSGYDAMOLRDR  
 MGRIMVWRHAKKEIYLVKEEPTSPFMKRVWVRASLILPYEWMLL"  
 complement(2038. .3195)  
 /gene="Ecs1425"  
 /complement(2038. .3195)  
 /note="Ecs1425"  
 /note="similar to MDOC\_ECOLI g113025226sp|P75920 percent  
 identity 99 in 385 aa (conserved in E.coli K-12)"  
 /codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="glucans biosynthesis protein"  
 /protein\_id="BAB34850.1"  
 /db\_xref="GI:13360889"  
 /translation="NNPVAQNEYFLDSTIRAMLLMIGIPHHISLYTSSHTHWNSAEP  
 SLMTLIPDFDIHSFMQVEFVYISGFYSYMLFARYPLKMKWKNVERGVIMPLAIPLL  
 TLPOEIMAOYKNGKESMFGSLSLDYKNTLAMELLSHMFLVLVMTTLCVAFKRI  
 RNLNDSLOATYKNGKESMFGSLSLDYKNTLAMELLSHMFLVLVMTTLCVAFKRI  
 FLPELITGILALAFIEPTIKALPTTPSGRGTIAAALFAVAYLNLNQRGSGDAMWETES  
 VITVPLGIMAMVNVSPFGRIKALTPGASRTGYTNASLFTYLWNLPLTPFGAITTPHIT  
 TSNWLGFLCGLEFVNGIATILTEIHLRIPLKFLFSGKPVKRENDAPAR"  
 3588. .5123  
 /gene="Ecs1426"  
 /note="Ecs1426"  
 /note="similar to MDOC\_ECOLI g111787286 percent identity  
 99 in 511 aa (conserved in E.coli K-12)"  
 /codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="periplasmic glucans biosynthesis protein"  
 /protein\_id="BAB34849.1"  
 /db\_xref="GI:13360890"  
 /translation="MMKKNKLSAAVMLTLTYTSSNAFSIDVVAKQASLAGKGYEAK  
 SNLPSVFDPMKADVQOOLQFNHDKAYMNNLTKTPFKTEHQGAYEPPTPVKINETYAAR  
 VKRIYSDDYEFEGVNDHDKDVKDLGDFAGFPKLYVINSKDKDEIYSMAGSAFYVY  
 GAGQYVYGSARGLADITDLPAGEEPREKFEWTERKPKDKRLTIIYALDSDPATGAY  
 KPVWPGEDTYVDVOSKITLRKVKGLGVALPLTSMFLRCPNDSPPANNRPELHDSNG  
 LSIHANGNEIMRPLNPNRHLNLSFSFSMNPGQFGLQKRDSPREDDIDRDYRDS  
 AMVPRKGEWKGKSVLEIPTMDENTDNVAAVTPQDLPEPKEMNFKYITPTSDRES  
 KLHANEAMVQOVRTSTGQVQSNNLRCPDGTAFVVDFTGAEMKILDETPVYAGTS  
 IGDNEIYVESVTVNPVTKGWRLVKRVKDKAKKTEMRALVNAQDTSETMSYGLP  
 ANE"  
 5086. .7659  
 /gene="Ecs1427"  
 5086. .7659  
 /gene="Ecs1427"  
 /note="similar to membrane glycosyltransferase MOON\_ECOLI  
 g111787287 percent identity 100 in 847 aa but differs at  
 111-ter (conserved in E.coli K-12)"  
 /codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="membrane glycosyltransferase"  
 /protein\_id="BAB34850.1"  
 /db\_xref="GI:13360891"  
 /translation="MKPGATSLYPNNKTEYIDAMPIDASEKALRPRTDIRAVHQALDD  
 AEHRWAPEDSDPSQVARYRLKAMPDLSLADQGLIDDESDQDLKAMPEAKRSMFPD  
 PMRTVPYRFMRNLRGRDTPYVRLATKEEDESQKWRVGRIRRYILITLTAQVY  
 VATWMTKILPQAGALINPMQODLPEHQAQLPYMLQTSILLYLFLVPCWVSAG  
 EWTALMGTLQQLIDGDKSISASTYGDDEPLNPEHQAQLPITCNEDYNNRPAQIRAR  
 ESKVTKGNAKHRTDVLILDSYIPDICAQVOKAMMELIAEYGGGQFYRRRRRRRRTK  
 SGNIDDFCRWMSQSYVYVADLDPYVTDCLCGIVRLMEANFNAIGIIOSSPASGND  
 TLAYCAQFATRYVYLPFAGAHFMOLGESHGHNAILIRKVFIEICALAPYBEGSS  
 FLATSLSHDFEYVLAAMRRGQGVYLAIDPGSEELPNDILDKRDRCHGNLME  
 RLATVKGHNVYARVRLGVMSVYLAIDPMFMEASLTALOVNLAIDLPQVDFRDLRF  
 EYVWPMREBLALALASTYVNLFLPRILSLILILWCKGTGKYGSGFMYRTLSLILEVRS  
 VVAVRRLHLEFVAVSALAEVWVNSPQRDDSTSWGEAFKRHSQDLGLGVAVAG  
 VAMLDRLLEFVAVPVSLLISPPVSVISRRATVGRTKRWKLFLLIPEEYSPQVYDVL

IDRLRLNNRQSLDDGFMHAFVNSFNALATAMATAKHRSKYLEIARDRIVEALND  
 IPEELNRDLRVLLSDPYETMARLHRVWNSPERSISWVSTIEGIKLNPALRKPDAS  
 Q"

gene  
 CDS  
 7832..8059  
 /gene="ECS1428"  
 7832..8059  
 /gene="ECS1428"  
 /note="Identical to YCEK\_ECOLI g111787288 (Conserved in  
 E.coli K-12)"  
 /codon\_start=1  
 /transl\_table=1  
 /evidence=not\_experimental  
 /product="hypothetical protein"  
 /protein\_id="BAB34851.1"  
 /db\_xref="GI:13360892"  
 /translation="MRLIVSIWTLSCGSISRTIPGGHGNGYYPGVMDVDRDS  
 AMRVYTLIDLPESLVPFTLLPIDIHHPPE"  
 complement(8060..8437)  
 /gene="ECS1429"  
 complement(8060..8437)  
 /gene="ECS1429"  
 /note="Similar to MSYB\_ECOLI g111787289 percent identity  
 99 in 125 aa (Conserved in E.coli K-12)"  
 /codon\_start=1  
 /transl\_table=1  
 /evidence=not\_experimental  
 /product="acidic protein MsyB"  
 /protein\_id="BAB34852.1"  
 /db\_xref="GI:13360893"  
 /translation="MRTVATLEPAIDNAREEFLADNPGLDAEDANVQFPAQKYVLAQ  
 DGDIMQVEFFADEGEGECPLMLSGEAAQSVFDGVDLEIRQWQDEMTLHEMDGEG

Query Match 4.6%; Score 22; DB 1; Length 222605;  
 Best Local Similarity 100.0%; Pred. No. 0.55;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 TAATCGGACACCGGTAAACAGC 312  
 ||||||||||||||||||  
 Db 105175 TAATCGGACACCGGTAAACAGC 105196

RESULT 13  
 AP002557/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 AUTHORS  
 REFERENCE  
 JOURNAL  
 MEDLINE  
 MEDLINE  
 REFERENCE

AP002557 286485 bp DNA linear BCT 07-MAR-2001  
 Escherichia coli O157:H7 DNA, complete genome, section 8/20.  
 AP002557 BA000007  
 AP002557.1 GI:13361466  
 Escherichia coli O157:H7 (strain:O157:H7, sub\_strain:RIMD 0509952)  
 DNA.  
 Escherichia coli O157:H7  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 1 (sites)  
 Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,  
 Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,  
 Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,  
 Sasaki,K., and Shinagawa,H.  
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the  
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7  
 derived from the Sakai outbreak  
 Genes Genet. Syst. 74 (5), 227-239 (1999)  
 20198780  
 2 (sites)  
 Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,  
 Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and  
 Hayashi,T.  
 Comparative analysis of the whole set of rRNA operons between an  
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an  
 Escherichia coli K-12 strain MG1655  
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
 20557356  
 3 (sites)

```

98 in 879 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB35426.1"
/db_xref="GI:13361469"
/translation="MLGKYKVALALLIIILVPLLLMTLGIWPTLAGIILPIG
ALDESPRTTRKGLIIIPDLRYLVGDCOLAHITNASHSRSLNGLVGEIISL
POTEEPSAPKPTIAQOMQSMIPNRIINDILSPQEMOQGLSTALSDIOOLRT
KYEFGQGLGOOLFSLDLYVARENDSPKVLAGEFTTLPMDGGLVGSCHATTALN
EPISTADALDWMQENSGSLIYLARDNDGPIIARDNDPITLQOOLYDGSCHATTALN
GRLEKVDNMQAGIENGLVSNLVSGRSLVITGOAGKGNALVIEPQSLSDMSNDQPL
EAKQADRIILYARIPADLGSLSDPITLTFEPGALLRSGRYVLSIDIDEIRMPJAL
TORVEDRILIAITQAHENLGDVFLMDGLANDFLPRAGQMRKRWKGSFTPMNA
DVAGCKMHSTITLTDLSITGFDQLOCTMTVEKPRILIKRFLVWRDAQHSFEGS
SLANGOTLITFGGSLVLEPSTLTKSVYDGRDPTLFLKGLIHAGEIVRVNGMDSGIR
GNAMRKQSLTYVFOPLVPPDKMKSNRDELEYLAQVARSAAEQDFRAGGCHVKGSS
MPLNQNQVDFVLPRFADGAWHIGTRGTPPTLRLAEVTLNIAKNTADLOGRYPM
EEELLLITDYSVLVGLGNVLMOILRMPORHDLRLNNLSSESLVSAVNPQFANSG
SGSLPLPTLNENKVIKQGMILKGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN
RSGATRLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN
RSLTRILNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN
GLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN
WLEQNALDEFTGCPGSGKCEKQ"
gene
4660 .4845
/gene="RCS3004"
4660 .4845
/contig="RCS3004"
/contig="RCS3004"

```

source

1. 286485

Location/Qualifiers

/organism="Escherichia coli O157:H7"

/strain="O157:H7"

/sub\_strain="RIMD\_050952"

/db\_xref="taxon:83334"

complement(294..716)

/gene="Ecs2001"

complement(294..716)

/gene="Ecs2001"

/note="similar to HSLJ\_ECOLI g11787644.percent identity 99 in 140 aa (Conserved in E.coli K-12)"

/codon\_start=1

/transl\_table=1

/evidence="not\_experimental"

/product="heat shock protein hslJ"

/protein\_id="BAB35424.1"

/db\_xref="GI:13361467"

/translation="MKRYAFALSLIMACVSNDRKIAYTPDQLHRRFVLESNGKPTDCKNPEISFEGEKMGISGMCNRFGEKLSNGELTKKGLAMRMAMCAPOLNEIDVTISELKGCAQVDTLANGLMGLATAKOTLVYKILADLM"

complement(827..1816)

/gene="Ecs2002"

complement(827..1816)

/gene="Ecs2002"

/note="similar to LDHA\_ECOLI g11787645.percent identity 99 in 329 aa (Conserved in E.coli K-12)"

/codon\_start=1

/transl\_table=1

/evidence="not\_experimental"

/product="fermentative D-lactate dehydrogenase"

/protein\_id="BAB35425.1"

/db\_xref="GI:13361468"

/translation="MKLAVYSTKQDKKYLDOOVNESGEPLFEFDLITTEKTAKTANG EAVAEHIGMMTLNRIHRAVORTPRANSLSEGELFTMYRGKTAGVIGTIGVAMT AFILGEMRLAFDEPYSAAALCEGVVDPLTFLSESDVISLHCLPIYENYHLINLA ODDVFRLSACHNVFTGHQAFILAEALNISISOTYLNLSNLEKGETCPNELV"

2024..4663

/gene="Ecs2003"

2024..4663

/gene="Ecs2003"

/note="similar to YDBH\_ECOLI g11787646.percent identity

gene

CDS

gene

CDS

gene

CDS

Gene  
CDS  
GKWLRR"  
5590..8625  
/gene="ECS2006"  
5590..8625  
/gene="ECS2006"  
/note="unknown, similar to Biga protein [Salmonella typhi]  
ded1AF133185-1|AD29677.1 percent identity 32 in 106 aa,  
similar to N-terminal part (1-1128 in 2020 aa) of YdbA.2  
[Escherichia coli] gi|7465766|pir|C48399 percent identity  
49 in 1011 aa"  
/codon\_start=1  
/trans\_table=11  
/evidence-not-experimental  
/product="putative Biga-like protein"  
/protein\_id="BAB35429.1"  
/db\_xref="gi:13361472"  
/translation="MQRKLISVCVMAALSSQTMAADSIIDSTKTRKSKITTCRAN  
IHLSAEQLENTASACREKNSAIVLPMAAGLAATATVLTSDDNHNHNSRVP  
DDGGDTPVPDDGGDTPVPDDGGDTPVPDDGGDTPVPDDGGDTPVPDDGGDTPVP  
DDGGDTPVPDDGGDTPVPDDGGDTPVPDDGGDTPVPDDGGDTPVPDDGGDTPVP  
KSMVTADSNIVTVDSGVNTSGGTLTMSYDDGGLIHTKEAGVVDGTTGMAIEGKATIV  
TTCGGNTALNGCTVMTDGNISLNDGKTTAIEGSGVCLTIGDNIITINNNGTEVD

JOURNAL MEDLINE REFERENCE AUTHORS	TITLE	GCTAVIINGDNTKLTACGSTITNGCTGSLINDGNARVDNOCMTSVYDGENSTGSKIVG DCATIKOEGLVIDGNDTPEFVNSKNGKINIVIEDNSIGMLDGDGVIVIMG DLNVGOAAGENAIGIOIDGNATFPVNDISATNAGTCVSVAGDKANISLAGIIVG DRESTDLDVSGNNNNMTLATELVAVTQKATGVVNSGDNGTIEIAGSILVDKQADNA OPIEFNPSTGVNISGNDNVTLTGOLTVVADSKTTSRSYASVDAQAEHLAGIYAADDD MPTLNGSTGVHVEKRVMDGSKPSARRIGDTPILINWDGSHPYLNGESTVIGSEFP LGEFNLIOJSHGALEIIGADATFMDSDVDSFTYRYVALSTIISDGAKATNGEVEL KNIGFAAMNKKDSTVINNGSIGLAMYDFGDDPAKPKVEDYEGGICGVNNGMPTKMMNO HSYLVNPAEMNLSDGTSFNKALGLTGLMALSYSSSTLNGCTGIDMAGRSVGMALID KSTADNEGQITLDTLWVDENDSTLSNNVANSTADFGVMASTGATYVNCALTKATAT NHENGVITVYNAGMAAYGNTMTVINOGTINLEKNENYDSIGVNRKLVGMAYVHEGT AINDOTGVNINAEINGOAFYNDGTGLIINVTCTGFCGVC"	
			8881. .11304 /gene="ECs2007" 8881. .11304 /gene="ECs2007" /note="unknown, similar to BigB protein [Salmonella typhi] dadlAF13185-21AD29678.1.percent identity 72 in 740 aa, similar to C-terminal part (1220-2020 in 2020 aa) of ydbA.2 [Escherichia coli] g117465766 p1r  C48399 percent identity 77 in 806 aa"
			/codon_start=1 /transl_table=11 /evidence=not_experimental
Query Match	4.6%; Score 22; DB 1; Length 286485;		
Best Local Similarity	100.0%; Pred. No. 0.56;		
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Oy 291	TAATCCGACACCGGTAAACAC 312		
Db 246743	TAATCCGACACCGGTAAACAC 246722		
RESULT 14			
AP002556	291136 bp DNA linear BCT 07-MAR-2001		
LOCUS	Escherichia coli O157:H7 DNA, complete genome, section 7/20.		
DEFINITION	AP002556 BA000007		
ACCESSION	AP002556.1 GI:13361156		
VERSION			
KEYWORDS			
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952) DNA.		
ORGANISM	Escherichia coli O157:H7		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
AUTHORS	1 (sites) Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,T., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak Genet. Syst. 74 (5), 227-239 (1999) 20198780		
TITLE	2 (sites) Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T. Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)		
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS	3 (sites) Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak		
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
COMMENT			
FEATURES			
SOURCE			
gene	Gene 258 (1-2), 127-139 (2000) 20564182 4 (sites) Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H. Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12 DNA Res. 8 (1), 11-22 (2001) 21156231 5 (bases 1 to 291136) Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T. Direct Submission Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047) genome project.		
CDS	location/Qualifiers 1. .291136 /organism="Escherichia coli O157:H7" /strain="O157:H7" /sub_strain="RIMD 0509952" /db_xref="taxon:83334" 126. .380 /gene="ECs1692" 126. .380 /gene="ECs1692" 126. .380 /note="similar to YWGE_ECOLI g111787445 percent identity 97 in 84 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB35115.1" /db_xref="GI:13361157" /translation="MGIIAMIFGLIAGIIAKLIMPGRDGGFFLTCLIGIVAVVGG WLATMFIGIGSIGFNLHSLFVAVGAILVLGVFLRRR" complement(430. .2400) /gene="ECs1693" complement(430. .2400) /note="probable TonB dependent outer membrane receptor, similar to TonB dependent outer membrane receptor Prta [Escherichia coli CFT073] g113661477 gb IAAC61709.1  percent identity 97 in 656 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative TonB dependent outer membrane receptor" /protein_id="BAB35116.1" /db_xref="GI:13361158" /translation="MRUKHYLCTALSLAFTQQAVALAQSDESLTWSSPVSSTTVLV DDPTKRALDKQNAQALSVYGVVGLDQSGSRNEQYKVRFSRQVPVFDGVPVIVP YDQNDLRLTLNNLDAVEYSKSSLLQGPQMGGAINTTQKPTKPLEASLGYRGG WRSQONAAVDMHSAFASSELGYLQVSSQQLKQDFLPGVANNIDLAGHKMINSR DKRGIVKIGFPTRENDYELIYIKODEKDNPPYSGNSQKSRWQPEYDKSFY OCCTQNDPRTLSRLYRDTFETLMNYSIDLKKKCSYSYSDYSDGAGIQIAD VRENDLITFAVNNKDVHREKGAHPAAYIDRETRVSLASEYQMAAADNVAGISTY DWRDSEVAKHHEKDSITHYDDNNSAFNMQVQKHFVNEDTLALSYDRKREPTLK ERYTTSKPAYNQIAIVNPOLKPERAGVDITWNGATFHDWGEFVSYYNRSAPAILSH NIDAPRTIONSGTVDYSGLDAGIKKISNIIDVGLSYALIHADAKRRDIDKLTDLPT QPTWAMTTLKPEWPLSVTLSEERASSYSNDSQKAAAGPAVTHIRADYTLTGHSFVN ASVNNLPTDIOYASGFEDEGRNPAAGVETP" complement(2426. .3280) /gene="ECs1694" complement(2426. .3280) /note="molybdenum transport protein, similar to molybdenum		

```

transport proteins e.g. g1|3661478|gb|AAC61710.1| percent
identity 91 in 284 aa
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="molybdenum transport protein"
/protein_id="BAB35117.1"
/db_xref="GI:13361159"
/translation="MIFLSQADIALLEDIQSGDLTTFRALNIGHQGYIEFLRQGG
CQGVSVACMKLTTLGITIDAVSAGSQAAGORLIRAGQNAALHQKKAIONVLEW
LLEFANRRHLNDQMSGAILNQLSRHNAPEKRIIVVADTPKEAIIARQPDVLQDKF
SLQAPASST"
complement(3219..3500)
/gene="ECS1695"
complement(3219..3500)
/gene="ECS1695"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB35118.1"
/db_xref="GI:13361160"
/translation="MPTLLSYDKTANRITAPRVLSRTSAGVLAIRLTMSVSGYIAG
INNMKLVPSPPLPAATGSGGIAVRRRCHDFPSSGYNRCVTSGRYPGR"
complement(3445..4089)
/gene="ECS1696"
complement(3445..4089)
/gene="ECS1696"
/note="unknown, similar to hypothetical proteins e.g. Orf2
|Escherichia coli CPT073| g1|3661479|gb|AAC61711.1|
percent identity 98 in 214 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB35119.1"
/db_xref="GI:13361161"
/translation="MLIDIDIRADLYLQOLKLAHRTKTPDHMPORAEKMAENCASPT
ASYLOOLIKITIDQGNQTLFDMGCGPQVSLADLKLTITGYDYSQGLNVARAA
ALKADNVHLIORMMEDMSDLPRCDIAVASRSTLVADROMKSLNNOALRYTTHL
VSTSPFSPAIORAGREVELEPNYFIALNVLYOMGIYAHVDFTRGNCQODNST"
complement(4099..4857)
/gene="ECS1697"
complement(4099..4857)
/gene="ECS1697"
/note="probable ferric enterobactin transport, similar to
ferric enterobactin transport ATP-binding protein
|Escherichia coli CPT073| g1|3661480|gb|AAC61712.1|
percent identity 79 in 148 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative ferric enterobactin transport
ATP-binding protein"
/protein_id="BAB35120.1"
/db_xref="GI:13361162"
/translation="MTLLAVRHASLGSYRHPVLDVSPTLSOGTICCLGANGCQYTT
LMRSIGVILDKGEVLDSLPVOTLSHROAQAIAVPOAHGQIRFESVLDVWMLG
LILDPASSLDLGHQIOLIDLTAOLKNGCKMTSHHPLHANAIDSIIQVEPDGRV
TQGLPTEDLITKLAALRYVSADQIHHLHLSAISH"
complement(4854..5834)
/gene="ECS1698"
complement(4854..5834)
/gene="ECS1698"
/note="probable ABC transporter protein (permease), similar
to ABC transporter permeases e.g. |Haemophilus influenzae|
aa|2501391|sp|Q57130|YE1_HAEIN percent identity 40 in 323
/codon_start=1

```

```

Query Match
Best Local Similarity 100.0%; Score 22; DB 1; Length 291136;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 291 TAATCCGACACCGGTACACGC 312
DB 101095 TAATCCGACACCGGTACACGC 101116

RESULT 15
AGHGBEHA/c
LOCUS
DEFINITION
Agoseris heterophylla partial 18S ribosomal RNA (18S rRNA) and 25S
ribosomal RNA (25S rRNA) genes; complete 5.8S ribosomal RNA (5.8S
rRNA) gene; complete internal transcribed spacers 1 and 2.
VERSION
L133954
KEYWORDS
18S ribosomal RNA; 25S ribosomal RNA; 5.8S ribosomal RNA.
SOURCE
Agoseris heterophylla
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Agoseris.
REFERENCE
1 (bases 1 to 702)
Kim, K.-J. and Jansen, R.K.
Comparison of phylogenetic hypotheses among different data sets in
dwarf dandelions (Krigia): additional information from internal
transcribed spacer sequences of nuclear ribosomal DNA
Plant Syst. Evol. (1993) In press
FEATURES
Location/Qualifiers
1..702
/organism="Agoseris heterophylla"
/db_xref="taxon:19017"
<1..32
/gene="18S rRNA"
/product="18S ribosomal RNA"
1..32
/gene="18S rRNA"
33..278
/product="internal transcribed spacer 1 (ITS1)"
279..442
/gene="5.8S rRNA"
/product="5.8S ribosomal RNA"
279..442
/gene="5.8S ribosomal RNA"
443..664
/gene="5.8S rRNA"
/product="internal transcribed spacer 2 (ITS2)"
665..2702
/gene="25S rRNA"
/product="25S ribosomal RNA"
665..702
/gene="25S rRNA"
BASE COUNT
151 a 187 c 174 t 1 others
ORIGIN

```

Query Match 4.3%; Score 21; DB 8; Length 702;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 GCGTGCCCTCAGCCGATGGC 456  
|||||  
Db 425 GCGTGCCCTCAGCCGATGGC 405

Search completed: October 27, 2002, 21:52:56  
Job time : 1195.58 secs

---

**THIS PAGE BLANK (USPTO)**





## RESULT 2

```

ID 054381 PRELIMINARY: PRT: 148 AA.
AC 054381;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DE 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE 17 KDA COMMON-ANTIGEN (FRAGMENT).
OC Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN NCBI_Taxid=789;
RP SEQUENCE FROM N.A.
RX MEDLINE=98087556; PubMed=9425244;
RT Davis M.J., Ying Z., Brunner B.R., Pantofa A., Ferwerda F.H.;
RL Rickettsial relative associated with papaya bunchy top disease.;
DR Curr. Microbiol. 36:80-84(1998).
FT NON_TER 1 1
FT NON_TER 1 1
SQ SEQUENCE 148 AA; 15050 MW; A7AFEEDE0AE8E4C CRC64;

Query Match
Best Local Similarity 40.18; Score 303.5; DB 2; Length 148;
Matches 55; Conservative 30; Mismatches 43; Indels 9; Gaps 2;

OY 25 SROEVGAATGAVGVAGAGLFGKSGRVSMATGAGLVGLGSGITGSGMDQDK---IK 80
DB 17 NKQSGSLIGTGLGVLGSGFGGGRILAAGAGLIGALIGNOIGAGMEDQRIKLELT 76
OY 81 LNSOLEKRVAGQVTRMRNDPTGNSYSVEPVYTYQRYNKREROQYCEEFQKAMIAQKQ 140
DB 77 SQRLAEAAAGSSGVYWRNDNGNGTVPKAY-----KNTGQYRETYTIVVGGKQ 131
OY 141 EITGTACPODPRMOYI 157
DB 132 KAVGTACRQPDGQWQV 148

```

## RESULT 3

```

ID 09F9F2 PRELIMINARY: PRT: 159 AA.
AC 09F9F2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DE 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 17 KDA GENUS-COMMON ANTIGEN.
OC Rickettsia felis (Rickettsia azadi).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN NCBI_Taxid=42862;
RP SEQUENCE FROM N.A.
RX MEDLINE=21217364; PubMed=11321078;
RT Bouyer D.H., Stenos J., Croquet-Valdes P., Moron C.G., Popov V.L.,
RA Zavala-Velazquez J.E., Foll L.D., Storchard D.R., Azad A.F.;
RT "Rickettsia felis: molecular characterization of a new member of the
RT spotted fever group.";
RT Int. J. Syst. Evol. Microbiol. 51:339-347(2001).
DR EMBL: AF195118; AAG28452.1; -
SQ SEQUENCE 159 AA; 16497 MW; 34C5B020AF470A1F CRC64;

Query Match
Best Local Similarity 37.98; Score 283; DB 2; Length 159;
Matches 61; Conservative 30; Mismatches 54; Indels 16; Gaps 6;

OY 8 SLLIITSV---FLVGC--AQNFSRQEVGAATGAVGVAGQLFGKSGRVSMATG--GAVL 61
DB 5 SKIMILIAASMLQACNCPGGMNKGIGTLLGAGGALLGSGQFGKQGL--VGCVGALL 63

```

## RESULT 4

```

ID 052252 PRELIMINARY: PRT: 137 AA.
AC 052252;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DE 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
OC Rickettsia coolleyi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN NCBI_Taxid=69410;
RP SEQUENCE FROM N.A.
RX Billings A.N., Teltow G.J., Walker D.H.;
RT "Molecular characterization of a novel spotted fever group rickettsial
RT species from ixodes scapularis in Texas.";
DR EMBL: AF031534; AAB95267.1; -
FT NON_TER 1 1
FT NON_TER 1 1
SQ SEQUENCE 137 AA; 14215 MW; A27597A9ADF85FC3 CRC64;

Query Match
Best Local Similarity 31.18; Score 258.5; DB 2; Length 137;
Matches 53; Conservative 25; Mismatches 46; Indels 11; Gaps 4;

OY 22 QNFSROEVGAATGAVGVAGAGLFGKSGRVSMATG--GAVLGLGSGITGSGMDQDK-- 78
DB 7 RGMKQGTGTLGGAGGALLGSGQFGKQGL--VGVGALLGAVLGGQIGAGMEDQRL 65
OY 79 --IKLNSLEKRVAGQVTRMRNDPTGNSYSVEPVYTYQRYNKREROQYCEEFQKAMIA 136
DB 66 AELTSORALEAPSSSTMRNDNGNGTVPKTY-----RNSGTQYRETYTIVVIG 120
OY 137 GORQETYGTAQPD 151
DB 121 GKOQKAYGNACRQPD 135

```

## RESULT 5

```

ID 09K2N6 PRELIMINARY: PRT: 144 AA.
AC 09K2N6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DE 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 17 KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN NCBI_Taxid=38028;
RP SEQUENCE FROM N.A.
RX Schuilenburg H.J.G.V.D., Habig M., Sloggett J.J., Webberly M.K.,
RT Bertrand D., Hurst G.D.D., Majerus M.E.N.;
RT "On the evolution of male-killing: Monophyletic origin and horizontal
RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-
RT generic ladybirds, Adalia bipunctata L. and A. decempunctata L.
RT (Coleoptera: Coccinellidae).";
RT Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ269518; CAB96383.1; -
FT NON_TER 1 1
FT NON_TER 144 144

```

```

DE 17 KDA ANTIGEN (FRAGMENT)..
OS male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=120393;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20575219; PubMed=11133455;
RA Schulenburg H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
RT Bertland D., Hurst G.D.D., Majerus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT Coccinellidae)".
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL; AJ269516; CAB96381.1; -.
FT NON_TER 1 1
FT NON_TER 144 144
SO SEQUENCE 144 AA; 14801 MM; CB25472F16A56AE7 CRC64;

Query Match 30.2%; Score 251.5; DB 2; Length 144;
Best Local Similarity 39.4%; Pred. No. 4.1e-15;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4

QY 25 SROEVGATGAVVGGVAGOLFGKSGRVSMAIG-GAVIGGLGSKRGOSMDOODK----I 79
Db 17 NKQGTGTLKGAGGALLGSGFGKRGQL-VGVGVGALLGAVGGGAGCAGMDQDRRLAEL 75
QY 80 KLNLSLEKVAQGVWRWNPDTGNSYSVEPVTRYRNKQERROOYCREFOOKAMTAGOR 139
Db 76 TSGRLLEAPSSSNVEMWNPDMGNNGYVTPNNTY-----RNSTGVCRCRYTQTVVIGKQ 130
QY 140 QEIVGTACPPD 151
Db 131 QKSYGNACRPD 142

RESULT 8
ID Q53154 PRELIMINARY; PRT; 154 AA.
AC Q53154;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE (CLONE PRB F1SF 1), 5' END CDS (FRAGMENT).
OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=789;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93084757; PubMed=1452660;
RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
RT "Characterization and comparison of Australian human spotted fever
RT group Rickettsiae.".
RL J. Clin. Microbiol. 30:2896-2902(1992).
DR EMBL; M9391; AAA73386.1; -.
FT NON_TER 154 154
SO SEQUENCE 154 AA; 15849 MM; F5C35855EDBA39D2 CRC64;

Query Match 29.3%; Score 244; DB 2; Length 154;
Best Local Similarity 35.3%; Pred. No. 2.1e-14;
Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6

QY 8 SLLITSY---FLVGC--AQNFSROEVGAATGAVVGGVAGOLFGRKSGRVSMAIG-GAVL 61
Db 5 SKIMITALATSMLOACNCPGGCNKKGCTGTLGAGAGALLGSGFGKRGQL-VGVGALL 63
QY 62 GLGISKRGOSMDOODK----IKLNLSLEKVAQGVWRWNPDTGNSYSVEPVTRYRN 117
Db 64 GAVIGGQIGAGMDQDRRLAELTSGRALETARSSNVEMWNPDMGNSYGVVTPNKTYNST 123
QY 118 KOERROOYCREFOOKAMTAGOKOELTYGTRACPPDGR 153

```

DB 124 GOD-----CRVYTOTVIGGKQOKAKYGNACRQPDGQ 154

## RESULT 9

Q9F909 PRELIMINARY; PRT; 151 AA.  
 AC Q9F909;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE OUTER MEMBRANE PROTEIN (FRAGMENT).  
 OS Rickettsia helvetica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=35789;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nilsson K., Pahlson C.;  
 RT "Novel peptide diagnostic reagent and kit for detection of  
 RL rickettsiosis.";  
 DT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 FT EMBL: AF181036; AAC09427.1; -  
 FT NON\_TER 151  
 SO SEQUENCE 151 AA; 15621 MW; B7407B9C71E4B39 CRC64;

Query Match 28.4%; Score 239; DB 2; Length 151;  
 Best Local Similarity 35.9%; Pred. No. 5.7e-14;  
 Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

QY 8 SLLIIIV---FLVGC--AQNFSROEVGAATGAVVGVAGQLFGKSGRVSMAIG-GAVL 61  
 DB 5 SKIMIIAASMLQACNGPGMKNKOGTGLLGAGGALLGSQFGKGGQL-VGVGALL 63  
 QY 62 GGLIGSTIGSMDOOK---IKLNSLEKAKQVTRMRNPDGTGNSYVEPRYQRYN 117  
 DB 64 GAVLGQOIVAGMDEORRLAELTSQLALEAPSSNVMENPNNGVYVTPKTY----- 119  
 QY 118 KQRROOYCRFOOKAMKAGKQKQEIYGTACPOP 150  
 DB 120 -RNSTGYCREYTOTVIGGKQOKAKYGNACRQPDGQ 151  
 SQ SEQUENCE 151 AA; 15621 MW; B7407B9C71E4B39 CRC64;

## RESULT 10

Q9F001 PRELIMINARY; PRT; 131 AA.  
 AC Q9F001;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 17 KDA PROTEIN (FRAGMENT).  
 OS Rickettsia sp. California 2.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=147259;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Roux V., Raoult D.;  
 RT "A new SFG rickettsia isolated from fleas."  
 DT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 FT [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Roux V., Raoult D.;  
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 FT EMBL: AF210693; AAG48554.1; -  
 FT NON\_TER 131  
 SO SEQUENCE 131 AA; 13374 MW; 23C8819E29FF860 CRC64;

Query Match 28.4%; Score 236.5; DB 2; Length 131;  
 Best Local Similarity 38.3%; Pred. No. 8e-14;  
 Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

QY 25 SROEVGAATGAVVGVAGQLFGKSGRVSMAIG-GAVLGGLIGSTIGSMDOOK-----I 79  
 DB 10 NKQGTGTLTGAGGALLGSQFGKGGQL-VGVGALLGAVLGQIGAGMDEORRLAEL 68  
 QY 80 KLNOSLEKAKQVTRMRNPDGTGNSYVEPRYQRYNKKQRRROOYCRFOOKAMKAGK 139  
 DB 69 TSQLALEAPSSNVMENPNNGVYVTPKTY-----RNSTGYCREYTOTVIGGKQ 123  
 QY 140 QEIYGTAC 147  
 DB 124 OKAYGNAC 131  
 SQ SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

## RESULT 11

Q52637 PRELIMINARY; PRT; 131 AA.  
 AC Q52637;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 17 KDA ANTIGEN (FRAGMENT).  
 OS Rickettsia sp.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=789;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94117373; PubMed=8288533;  
 RA Herren J.H., Hurst G.D., Zhang W., Breuer J.A., Stouthamer R.,  
 RT "Rickettsial relative associated with male killing in the ladybird  
 RL beetle (Adalia bipunctata)."  
 DT EMBL: U04162; AAA19235.1; -  
 FT NON\_TER 131  
 FT NON\_TER 131  
 SQ SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

Query Match 28.4%; Score 236.5; DB 2; Length 131;  
 Best Local Similarity 38.3%; Pred. No. 8e-14;  
 Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

QY 25 SROEVGAATGAVVGVAGQLFGKSGRVSMAIG-GAVLGGLIGSTIGSMDOOK-----I 79  
 DB 10 NKQGTGTLTGAGGALLGSQFGKGGQL-VGVGALLGAVLGQIGAGMDEORRLAEL 68  
 QY 80 KLNOSLEKAKQVTRMRNPDGTGNSYVEPRYQRYNKKQRRROOYCRFOOKAMKAGK 139  
 DB 69 TSQLALEAPSSNVMENPNNGVYVTPKTY-----RNSTGYCREYTOTVIGGKQ 123  
 QY 140 QEIYGTAC 147  
 DB 124 OKAYGNAC 131  
 SQ SEQUENCE 131 AA; 13344 MW; A1DCF71050DF52DF CRC64;

## RESULT 12

Q9L522 PRELIMINARY; PRT; 131 AA.  
 AC Q9L522;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 17 KDA SURFACE ANTIGEN (FRAGMENT).  
 OS Rickettsia peacockii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=47589;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=DAE1007;  
 MEDLINE=21091941; PubMed=1157215;  
 RX Simser J.A., Palmer A.T., Munderloh U.G., Kurtli T.J.;

RT "Isolation of a spotted fever group rickettsia, Rickettsia peacockii,  
in a Rocky Mountain wood tick, Dermacentor andersoni, cell line."  
RT Appl. Environ. Microbiol. 67:546-552(2001).

DR EMBL: AF260571; AAF69012.1; -.

FT NON\_TER 1 1  
SEQUENCE 131 AA; 13413 MW; 228C020550CAA9D0 CRC64;

Query Match 28.3%; Score 235.5; DB 2; Length 131;  
Best Local Similarity 38.3%; Pred. No. 9.8e-14;  
Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

OY 25 SROEVGATGAVYGVAGQLFGKSGRVSMAIG-GAVLGGLIGSKIGOSMDQDK---I 79  
Db 10 NKQGTGGLGAGGALLGSGQFGKSGQL-VGVGALLGAVLGQIGAGMDEQDRRLAE 68

OY 80 KLNQSLKRVKAGQYTRNRPDGTGNSYSVEPYRYQRYNKOERROQYCREFOOKAMIAQK 139  
Db 69 TSGRALETAPSGSNVWRPNDCNGYGVTPNKTY-----RNSTGYCREYTOTVIGGKQ 123

OY 140 QETYGTC 147  
Db 124 OKAYGNAC 131

## RESULT 13

ID 031208 PRELIMINARY; PRT; 105 AA.

AC 031208;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE 17 KDA ANTIGEN (FRAGMENT);

OS Rickettsia sp. 'La Copita';

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_TaxID=69475;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LA COPITA;

RX MEDLINE=98367252; Pubmed=9701930;

RA Billings A.N., Yu X.J., Teel P.D., Walker D.H.;

RT "Detection of a spotted fever group rickettsia in Amblyomma cajennense

(Acari: Ixodidae) in south Texas."

RL J. Med. Entomol. 35:474-478(1998).

DR EMBL: AF03499; AAB86943.1; -.

FT NON\_TER 1 1

FT NON\_TER 105 105

SO SEQUENCE 105 AA; 11236 MW; 7BD035104701D4FB CRC64;

Query Match 26.0%; Score 216.5; DB 2; Length 105;  
Best Local Similarity 40.4%; Pred. No. 3.7e-12;  
Matches 46; Conservative 17; Mismatches 34; Indels 17; Gaps 3;

OY 42 GOLFSGKSGRVSMAIGAVLGGLIGSKIGOSMDQDK---IKLNQSLKRVKAGQYTRN 97  
Db 3 GQLVGVV-----GALLGAVLGQIGAGMDEQDRRLAEISQRALETAPSGSNVWR 54

OY 98 NPDGTGNSYSVEPYRYQRYNKOERROQYCREFOOKAMIAQKQETYGTCPOP 151  
Db 55 NPDGNGYGVTPNKTY-----RNSTGYCREYTOTVIGGKQKAYGNACQOP 103

## RESULT 14

ID 09AGC7 PRELIMINARY; PRT; 77 AA.

AC 09AGC7;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE 17 KDA ANTIGEN (FRAGMENT);

OS Rickettsia typhi.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_TaxID=785;

RN [1]

RP SEQUENCE FROM N.A.

RA Paspaliaris V., Liedtke B., Viletta L., Whiting J.L.;

RT "Rickettsia typhus and Rickettsia felis rickettsioses in Australia."

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF336794; AKK21272.1; -.

FT NON\_TER 1 1

FT NON\_TER 77 77

SO SEQUENCE 77 AA; 7955 MW; B1E447C037263918 CRC64;

Query Match 16.7%; Score 139; DB 2; Length 77;  
Best Local Similarity 39.5%; Pred. No. 2.2e-05;  
Matches 30; Conservative 15; Mismatches 25; Indels 6; Gaps 3;

OY 42 GOLFSGKSGRVSMAIG-GAVLGGLIGSKIGOSMDQDK---IKLNQSLKRVKAGQYTR 96  
Db 3 GSGRGHGKGL-VGVGALLGAVLGQIGAGMDEQDRRLVELTSQRALETAPSGSNVWR 61

OY 97 RNPDTGNSYSVEPYRT 112  
Db 62 RNPDCNGHGYVTPNKTY 77

## RESULT 15

ID 0985G4 PRELIMINARY; PRT; 199 AA.

AC 0985G4;

DT 01-OCT-2001 (TREMBLrel. 18, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE MLR7687 PROTEIN.

GN MLR7687.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI\_TaxID=381;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MAFE303099;

RX MEDLINE=21082930; Pubmed=11214968;

RA Kano T., Nakamura Y., Sato S., Asami T., Kato T., Sasamoto S.,

RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimp S., Sugimoto M.,

RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

Mesorhizobium loti."

RL DNA Res. 7:331-336(2000).

DR EMBL: AP003012; BAB54098.1; -.

KW Complete proteome.

SO SEQUENCE 199 AA; 20517 MW; 98E6F20A734637AA CRC64;

Query Match 16.5%; Score 137; DB 16; Length 199;  
Best Local Similarity 30.9%; Pred. No. 0.00011;  
Matches 34; Conservative 16; Mismatches 50; Indels 10; Gaps 3;

OY 49 SGRVSMAIGAVLGGLIGSKIGOSMDQDKIKL-----NQSLEKRVKAGQYTRNRPDGTGNS 104  
Db 92 SGKVTKSLISMDSGLIGSGIGNGLSDPEKRSALAEYKKALEYTSQYVAMKGDQASHY 151

OY 105 YSVEPYRYQRYNKOERROQYCREFOOKAMIAQKQETYGTCPOPDCRW 154  
Db 152 GEVVPAPY-RVGSQD-----CROYTOTVTFGGAGVYTAGTACRNADGSW 195

Search completed: October 27, 2002, 10:59:53  
Job time : 19.0743 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 5.83938 Seconds

(without alignments)  
1067.533 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832  
Sequence: 1 MRCCLOGSSLIITSVLVGC.....ITGTACPDGKRWQVISTEK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	288	34.6	159	17KD_RICPR	P16624 rickettsia
2	286	34.4	159	17KD_RICJA	O52764 rickettsia
3	285	34.3	159	17KD_RICCN	P05372 rickettsia
4	276.5	33.2	159	17KD_RICTY	P22882 rickettsia
5	268	32.2	154	17KD_RICAU	P30928 rickettsia
6	259	31.1	154	17KD_RICPA	P50930 rickettsia
7	259	31.1	154	17KD_RICRH	P50931 rickettsia
8	256	30.8	154	17KD_RICMO	P50929 rickettsia
9	255	30.6	154	17KD_RICAM	P50927 rickettsia
10	156	18.8	80	17KD_RICCA	P29657 rickettsia
11	112.5	13.5	155	PCP_YEREN	P31464 yerstinia en
12	105.5	12.7	155	SLVB_SALTY	O33549 salmonella
13	104.5	12.6	155	SLVB_ECOLI	P55741 escherichia
14	101	12.1	155	PCP_HAEIN	P10325 haemophilus
15	100	12.0	179	YCFI_ECOLI	P37765 escherichia
16	90	10.8	179	KICJ_BOVIN	P66354 bos taurus
17	90	10.8	1332	KKDO_BACSU	P54334 bacillus su
18	88	10.6	1585	YQBO_BACSU	P45931 bacillus su
19	86.5	10.4	72	OSMB_SALTY	P37723 salmonella
20	85.5	10.3	72	OSMB_ECOLI	P17873 escherichia
21	84.5	10.2	431	KRE2_CANAL	O00310 candida alb
22	82.5	9.9	243	CYSK_SALTY	P17853 salmonella
23	82.5	9.9	541	NU37_YEAST	P48857 sacharomyc
24	82	9.9	132	V615_AQUAE	O66857 aquilex aeo
25	81.5	9.8	301	STXG_RAT	O92158 rattus norv
26	81.5	9.8	526	VPS_BTIV1	P33476 bluetongue
27	80	9.6	806	ITB7_MOUSE	P26011 mus musculi
28	79.5	9.6	113	KR33_CABEL	P34309 caenorhabdi
29	79.5	9.6	263	CANS_BOVIN	P13135 bos taurus
30	79	9.5	593	KICJ_HUMAN	P13645 homo sapien
31	78.5	9.4	219	VIAD_ECOLI	P37665 escherichia
32	78.5	9.4	747	SPD1_NEPCL	P19837 nephila cla
33	78	9.4	514	ATPA_THIFE	P11167 thiodacillu

34	77.5	9.3	359	1	ATPA_BOVIN	P19482 bos taurus
35	77.5	9.3	467	1	HML1_MYCLE	P46724 mycobacteri
36	77.5	9.3	543	1	ATPA_RAT	P15999 rattus norv
37	77.5	9.3	553	1	ATPA_HUMAN	P23705 homo sapien
38	77	9.3	266	1	CANS_RABIT	P06813 oryctolagus
39	76.5	9.2	553	1	ATPO_BOVIN	P19483 bos taurus
40	76.5	9.2	553	1	ATPA_MOUSE	O03265 mus musculi
41	76.5	9.2	559	1	KICJ_MOUSE	P02535 mus musculi
42	76.5	9.2	727	1	IF2M_HUMAN	P46189 homo sapien
43	76.5	9.2	747	1	EL5_BOVIN	P04985 bos taurus
44	76.5	9.2	2516	1	CCAD_DROME	O24270 drosophila
45	76	9.1	526	1	VPS_BTIV10	P07389 bluetongue

## ALIGNMENTS

```

RESULT 1
17KD_RICPR          STANDARD;          PRT:          159 AA.
AC  P16624;
DT  01-AUG-1990 (Rel. 15, Created)
DT  01-AUG-1991 (Rel. 19, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  17 kDa surface antigen precursor.
GN  OMP OR RP833.
OS  Rickettsia prowazekii.
OC  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC  Rickettsiaceae; Rickettsiae; Rickettsia.
OX  NCBI_Taxid=782;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MADRID E.
RX  MEDLINE=89359171; PubMed=2768201;
RA  Anderson B.E., Tzianabos T.;
RT  "Comparative sequence analysis of a genus-common rickettsial antigen
RT  gene.";
RL  J. Bacteriol. 171:5199-5201(1989).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MADRID E.
RX  MEDLINE=9039499; PubMed=9823893;
RA  Anderson S.G.E., Zomrodipour A., Andersson J.O.,
RA  Sichteritz-Ponten T., Alsmark U.C.M., Pedersson R.M., Naeslund A.K.,
RA  Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT  "The genome sequence of Rickettsia prowazekii and the origin of
RT  mitochondria.";
RL  Nature 396:133-140(1998).
CC  -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC  anchor (Probable).
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: M28482; AAA26378.1; ALT_SEQ.
DR  EMBL: AJ235273; CA15258.1; -.
DR  PIR: D33971; D33971.
DR  PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW  Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT  SIGNAL          1
FT  CHAIN           20      159      17 KDA SURFACE ANTIGEN.
FT  LIPID           20      20      N-ACYL DIGLYCERIDE (PROBABLE).
SQ  SEQUENCE       159 AA; 16672 MW; A33D404855EBB071 CRC64;
Query Match          34.6%; Score 288; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 3,1e-18;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

```

OY 8 SLLIISV---FLVGC--AQNFSROBVGATGAVGVAGOLFPGKSGRVSNAIG--GAVL 61  
 Db 5 SKIMIITALASMLQACNGSSGMKOGTGTLLGAGAGALLSGFGQKGL--VGVGALL 63  
 OY 62 GGLIGSKIGOSMDODK-----IKLNOSLEKAGAOVFRMRNPTGNSYVEVRYORYN 117  
 Db 64 GAVLGGQIGAGMDQDRRLLELTQSRLAESAPSGSNENRNDNGNHGVTYFNKTY----- 119  
 OY 118 KERRRQOYCGREFPOOKAMJAGOKOETITACPOPDGRMVOVIS 158  
 Db 120 -RNSAGYCRREYQTIVILGKQCKTKYGMACRQPDGQWQVYN 159

## RESULT 2

17KD\_RICJA  
 ID 17KD\_RICJA STANDARD; PRT; 159 AA.  
 AC Q52764;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 17 kDa surface antigen precursor.  
 GN OMP.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX NCBI\_TaxID=35790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YH.  
 RX MEDLINE=95229950; PubMed=7714214;  
 RA Furuya Y., Katayama T., Yoshida Y., Kaho T.;  
 RT "Specific amplification of Rickettsia japonica DNA from clinical  
 RL J. Clin. Microbiol. 33:487-489(1995).  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D16515; BAA03965.1;  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 159 BY SIMILARITY  
 FT LIPID 20 20 17 KDA SURFACE ANTIGEN.  
 SQ SEQUENCE 159 AA; 16554 MW; CDDCE7CEBDCD6B41 CRC64;  
 Query Match 34.4%; Score 286; DB 1; Length 159;  
 Best Local Similarity 37.9%; Pred. No. 4.6e-18;  
 Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROBVGATGAVGVAGOLFPGKSGRVSNAIG--GAVL 61  
 Db 5 SKIMIITALASMLQACNGSSGMKOGTGTLLGAGAGALLSGFGQKGL--VGVGALL 63  
 OY 62 GGLIGSKIGOSMDODK-----IKLNOSLEKAGAOVFRMRNPTGNSYVEVRYORYN 117  
 Db 64 GAVLGGQIGAGMDQDRRLLELTQSRLAESAPSGSNENRNDNGNHGVTYFNKTY----- 119  
 OY 118 KERRRQOYCGREFPOOKAMJAGOKOETITACPOPDGRMVOVIS 158  
 Db 120 -RNSAGYCRREYQTIVILGKQCKTKYGMACRQPDGQWQVYN 159

## RESULT 3

17KD\_RICCN  
 ID 17KD\_RICCN STANDARD; PRT; 159 AA.  
 AC Q52764;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 17 kDa surface antigen precursor.  
 GN OMP.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX NCBI\_TaxID=35790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YH.  
 RX MEDLINE=95229950; PubMed=7714214;  
 RA Furuya Y., Katayama T., Yoshida Y., Kaho T.;  
 RT "Specific amplification of Rickettsia japonica DNA from clinical  
 RL J. Clin. Microbiol. 33:487-489(1995).  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

AC P05372;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE 17 kDa surface antigen precursor.  
 GN OMP OR RCI287.  
 OS Rickettsia conorii, and  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OX NCBI\_TaxID=781, 783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-R.conorii, and R.rickettsii;  
 RX MEDLINE=89359171; PubMed=2768201;  
 RA Anderson B.E., Tzianabos T.;  
 RT "Comparative sequence analysis of a genus-common rickettsial antigen  
 RL J. Bacteriol. 171:5199-5201(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-R.conorii, and R.rickettsii;  
 RX MEDLINE=21442074; PubMed=1157893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
 RL Science 293:2093-2098(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-R.rickettsii;  
 RX MEDLINE=87222152; PubMed=3108232;  
 RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDade J.E.,  
 RA Fu Z.Y., Bellini W.J.;  
 RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia  
 RL J. Bacteriol. 169:2385-2390(1987).  
 RN [4]  
 RP SEQUENCE OF 1-30 FROM N.A.  
 RC SPECIES-R.rickettsii;  
 RX MEDLINE=8908059; PubMed=3139629;  
 RA Anderson B.E., Baumstark B.R., Bellini W.J.;  
 RT "Expression of the gene encoding the 17-kilodalton antigen from  
 RL Rickettsia rickettsii: transcription and posttranslational  
 RL J. Bacteriol. 170:4493-4500(1988).  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M28479; AAA26379.1;  
 DR EMBL; M28480; AAA26376.1;  
 DR EMBL; AE008675; AAL03825.1;  
 DR EMBL; M16486; AAA26381.1;  
 DR EMBL; J03371; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A25972; A25972.  
 DR PIR; A31836; A31836.  
 DR PIR; A33971; A33971.  
 DR PIR; B33971; B33971.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.  
 FT SIGNAL 1 19  
 FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT CONFLICT 146 146 N -> D (IN REF. 3).  
 FT CONFLICT 153 153 G -> E (IN REF. 3).



```

SQ SEQUENCE 159 AA; 16581 MW; 206A2BBF4FCE169 CRC64;
Query Match 34.3%; Score 285; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 5.6e-18;
Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;
OY 8 SLLIIIV---FLVGC--AQNFSRQEVGAATGAVVGVAGOLFPGKSGSVMAIG-GAVL 61
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 5 SKIMIIALATSMQLACNGPCGMNKGOTGTLGLGAGALLGSFGKKGQL-VGVGVALL 63
OY 62 GLLIISIKQSMDQDK---IKLMSLETKVAGQVTRRNNDPTGNSVEPEVRTYORYN 117
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 64 GAVLGQIGAGAGDEDDRLRAELTSGRALETAASGGSVNERNRNDNGNGYVTNKKY----- 119
OY 118 KOERROOYCREFOQKAMTAGOKOEIYGTACPOPDGRMOYIS 158
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 120 -RNSTGOYCREYQTQVTVIGGKQAKAYGNACRQPDQOMQVYN 159

RESULT 4
17KD_RICTY STANDARD: PRT; 159 AA.
AC P22882;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen gene."
RL J. Bacteriol. 171:5199-5201(1989).
CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce
CC or send an email to license@isb-sdb.ch).
-----
CC EMBL; M28481; AAA26377.1; -.
DR PIR; C33971; C33971.
DR PROSITE; PS00013; PROKAR.LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 159 AA; 16549 MW; 08973E2648F8DCD8 CRC64;

Query Match 33.2%; Score 276.5; DB 1; Length 159;
Best Local Similarity 39.6%; Pred. No. 3.1e-17;
Matches 55; Conservative 27; Mismatches 46; Indels 11; Gaps 4;
OY 25 SHQEVGAATGAVVGVAGOLFPGKSGRVSMAIG-GAVILGLLISKIGQMDQDK---I 79
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 27 NKQGTGTLGAGGALLGSQFGHGQL-VGVGVGALLAVLAGGQIGASLDEDDRLLLEL 85
OY 80 KLNQGLEKRVAGQVTRRNNDPTGNSVEPEVRTYORYNKOERROOYCREFOQKAMTAGOK 139
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 86 TSGRALESAPSGNSIEWRPNPDNGHGYVTPNKTY-----RNSTGOYCREYQTQVTVIGGQ 140
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

DB      141 QTYGNACRQPDGQWQVYN 159

RESULT 5
17KD_RICAU
ID      17KD_RICAU      STANDARD:      PRT:      154 AA.
AC      P50928:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
GN      OMP.
OS      Rickettsia australis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxID=787;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Baird R.W., Ross B., Dwyer B.
RL      Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
CC      - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      anchor (Probable).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib.ch).
-----
DR      EMBL: M74042; AAA6394.1;
DR      PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KM      Outer membrane; Lipoprotein; Antigen; Signal.
FT      SIGNAL 1 19 BY SIMILARITY.
FT      CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT      LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT      NON_TER 154 154
SQ      SEQUENCE 154 AA; 15967 MW; E3AA833346FAC320 CRC64;

Query Match 32.2%; Score 268; DB 1: Length 154;
Best Local Similarity 37.8%; Pred. No. 1.6e-16;
Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;

OY      8 SLLIIISV---FLVGCQAQ--NFSRQEVGAATGAVVGVAGQILFGKSGRVSMAIG-GAVL 61
DB      5 SKIMTIIALASLQKCNBSGCMNKGCTGTLGLGAGALLGSGFGKQGL-VGVEGALL 63
OY      62 GGLIGSKTIGQSDQDQDK---IKLNQSLSEKVKAGOVTRMRNPDGTGNSYSEVPRTYQRYN 117
DB      64 GAVLGGQIGAGDEDDRLRLAELTSGRALETAPSGNVNERNDNGNYGVPNKKTYRBSN 123
OY      118 KQERRQVCRERQOKAMLAGOKQELTYGACPPQDPR 153
DB      124 ----GQYCREYTORVIVIGKQOKAYGNACRQPDQ 154

RESULT 6
17KD_RICPA
ID      17KD_RICPA      STANDARD:      PRT:      154 AA.
AC      P50930:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
GN      OMP.
OS      Rickettsia parkeri.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxID=35792;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=MACULATUM.

```

RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;  
 CC Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U11020; AAB07706.1;  
 KW PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KM Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20  
 FT NON\_TER 154 154  
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5FEC CRC64.

Query Match  
 Best Local Similarity 36.5%; Score 259; DB 1; Length 154;  
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVGVAGQLFGKSGRVSMAIG-GAVL 61  
 DB 5 SKIMVIALAASMLQACNGPGMKNKOGTGLLGAGAGALLSGQFGKGGQL-VGVGVGALL 63  
 OY 62 GGLISKRGSGMSDQDK-----IKLNSLEKVKAGQVTRMRNDPTGNSYSVEPVRTYORYN 117  
 DB 64 GAVLGQIGAGMDEDDRLAELTSGRALETPASGSNVEMRNDNGNYGYTPNKTY----- 119  
 OY 118 KQRRQGYCREFOQKAMTAGOKOELTYGTACPOPDGR 153  
 DB 120 -RNSTGQYCREYTGTVVIGGKQKAKAGNACLOPDQ 154

RESULT 7  
 ID 17KD\_RICRH STANDARD; PRT; 154 AA.  
 AC P50931.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia rhipicephali.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_TaxID=33992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U11020; AAB07706.1;  
 KW PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KM Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20  
 SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5FEC CRC64.

FT NON\_TER 154 154  
 SQ SEQUENCE 154 AA; 15895 MW; 0CF85AD5D96DFEFB CRC64;

Query Match  
 Best Local Similarity 36.5%; Score 259; DB 1; Length 154;  
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVGVAGQLFGKSGRVSMAIG-GAVL 61  
 DB 5 SKIMVIALAASMLQACNGPGMKNKOGTGLLGAGAGALLSGQFGKGGQL-VGVGVGALL 63  
 OY 62 GGLISKRGSGMSDQDK-----IKLNSLEKVKAGQVTRMRNDPTGNSYSVEPVRTYORYN 117  
 DB 64 GAVLGQIGAGMDEDDRLAELTSGRALETPASGSNVEMRNDNGNYGYTPNKTY----- 119  
 OY 118 KQRRQGYCREFOQKAMTAGOKOELTYGTACPOPDGR 153  
 DB 120 -RNSTGQYCREYTGTVVIGGKQKAKAGNACLOPDQ 154

RESULT 8  
 ID 17KD\_RICMO STANDARD; PRT; 154 AA.  
 AC P50929.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia montana.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_TaxID=33991;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 CC anchor (Probable).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20  
 FT NON\_TER 154 154  
 SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E31DA CRC64;

Query Match  
 Best Local Similarity 36.5%; Score 256; DB 1; Length 154;  
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVGVAGQLFGKSGRVSMAIG-GAVL 61  
 DB 5 SKIMVIALAASMLQACNGPGMKNKOGTGLLGAGAGALLSGQFGKGGQL-VGVGVGALL 63  
 OY 62 GGLISKRGSGMSDQDK-----IKLNSLEKVKAGQVTRMRNDPTGNSYSVEPVRTYORYN 117  
 DB 64 GAVLGQIGAGMDEDDRLAELTSGRALETPASGSNVEMRNDNGNYGYTPNKTY----- 119  
 OY 118 KQRRQGYCREFOQKAMTAGOKOELTYGTACPOPDGR 153  
 DB 120 -RNSTGQYCREYTGTVVIGGKQKAKAGNACLOPDQ 154

```

RESULT 9
17KD_RICAM STANDARD: PRT; 154 AA.
AC P50927;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia amblyomml.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33989;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO 85-1084;
RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U11013; AAB07704.1; -
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15879 MW; E4FBE4C29D943581 CRC64;

Query Match 30.6%; Score 255; DB 1; Length 154;
Best Local Similarity 36.5%; Pred. No. 2,1e-15;
Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

QY 8 SSIITISV---FLVGC--AONFSROEVGATGAVGVGAGQLGKSGSRISMAIG-GAVL 61
DB 5 SKIMTIALASTLQACNGPCGMKOGTGLLGAGGALLGSGFGKGGQL-VGVGAGALL 63
QY 62 GGLIGSKIGSQMDQDK---IKLNSLEKVKAGGYTRMRNPDTGNSYSEVPRTYQRYN 117
DB 64 GAVLGQGVAGMGDEQDRRIARELTSOKALETAPRASNVENRNPNGNIGYTPPKTY- 119
QY 118 KQERRQOYCREFOQKAMIAQKQKQEIYGTACAPDPGR 153
DB 120 -RNSTGQYCREYQTYVIGGKQKQKAYGNACRDPDQ 154

RESULT 10
17KD_RICCA STANDARD: PRT; 80 AA.
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canad.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=92108069; Pubmed=1729713;

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M82879; -; NOT_ANNOTATED_CDS.
DR InterPro: IPR000437; Prok_lipoprot.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 156; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 3.9e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 42 GOLFGKSGRVSNAIGAVLGLIGSKRIGSQMDQDK---IKLNSLEKVKAGGYTRMR 97
DB 1 GSPFGKGGKGLIVGAGALLGATLGNIGAGMGEDQDRRLAELTSQRALETTPGTSIEWR 60
QY 98 NPDTGNSYSEVPRTYQ 114
DB 61 NPDNGNGYVTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD: PRT; 155 AA.
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1996 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcg precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RX MEDLINE=92121089; Pubmed=1732192;
RA Baeninger A.J., Hantke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli.";
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURUM SLTB AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X60448; CAA42977.1; -

```

```

DR   PIR: S23787, S23787.
DR   HSSP: P00778, 1P04.
DR   PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Outer membrane; Lipoprotein; signal.
FT   SIGNAL          1
FT   CHAIN           1
FT   LIPID           18
FT   LIPID           18
SQ   SEQUENCE       155 AA; 15362 MW; 8AD6BE2132E849FA CRC64;
                                N-ACYL DIGLYCERIDE (POTENTIAL).
Query Match
Best Local Similarity 24.7%; Score 112.5; DB 1; Length 155;
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4
OY   9 SLIIISVFLVGCANQ-----FSROE-----
DB   7 AVAIAAVLTLCANNNNTSLGDFVSASQAKQVYTYTGTLISVPRTYIGGDDNNWGAIG 33
OY   34 GAVGVGAVGQLEFGKSGKSGVSMATIGAVLGLGISKIGSQNDQDKTKL-----
DB   67 GAVIGFGELGNVTGGGCTGRSLATAGAGAGMGAGGVGAMNRTDGVLEVRKDDGTTILV 81
OY   82 --NQLSEKVKAGQVTPRRNPDTGNSVSP 109
DB   127 VQKGGPTRESVGO--RVMLASSGSTVTYVSP 154

RESULT 12
SLYB_SALTY
ID   SLYB_SALTY          STANDARD;          PRT;   155 AA.
AC   053549;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DE   01-MAR-2002 (Rel. 41, Last annotation update)
GN   SLYB OR STM1445 OR STY1677.
OS   Salmonella typhimurium, and
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella
ON   NCBI_TaxID=602, 601;
RX   1;
RC   SEQUENCE FROM N.A.
RC   MEDLINE=21539448; PubMed=8544813;
RX   MEDLINE=96133688; PubMed=8544813;
RA   Ludwig A.; Tenggel C.; Bauer S.; Bubert A.; Benz R.; Mollenkopf H.-J.;
RA   Goebel W.;
RT   "SLYA, a regulatory protein from Salmonella typhimurium, induces a
RT   haemolytic and pore-forming protein in Escherichia coli.";
RL   Mol. Gen. Genet. 249:474-486(1995).
[2]
RC   SEQUENCE FROM N.A.
RC   SPECIES=S.typhimurium; STRAIN=L72 / SCS1412 / ATCC 700720;
RX   MEDLINE=21539448; PubMed=11677609;
RA   McClelland M.; Sanderson K.E.; Spieth J.; Clifton S.W.; Latreille P.;
RA   Courtney L.; Porwollik S.; Ali J.; Dante M.; Du F.; Hou S.; Layman D.;
RA   Leonard S.; Nguyen C.; Scott K.; Holmes A.; Grewal N.; Mulvaney E.;
RA   Ryan E.; Sun H.; Florea L.; Miller W.; Miller W.; Stoneking T.; Nhan M.;
RA   Waterston R.; Wilson R.K.;
RT   Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
[3]
RC   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parthani J.; Dougan G.; James K.D.; Thomson N.R.; Pickard D.; Main J.;
RA   Churcher C.; Mungall K.L.; Bentley S.D.; Holden M.T.G.; Sebahia M.;
RA   Cronin A.; Davis P.; Davies K.M.; Chillingworth T.; Connor P.;
RA   Feltwell T.; Hanlin N.; Haque A.; Dowd L.; White N.; Farrar J.;
RA   Krogh A.; Larsen T.S.; Leather S.; Hien T.T.; Holroyd S.; Jajais K.;
RA   Quail M.; Rutherford K.; Leather S.; Moulse S.; O'Garra P.; Parry C.;
RA   Whitehead S.; Bartell B.G.;

```

```

RT      "Complete genome sequence of a multiple drug resistant Salmonella
RT      Enterica serovar Typhi CT18."
RL      Nature 413:848-852(2001).
CC      -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      anchor (Potential).
CC      -1- SIMILARITY: TO E.COLI SLTB, H-INFLUENZAE PCP AND Y. ENTEROCOLITICA
CC      PCP.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercia
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.slb.ch).
CC      -----
DR      EMBL; S80790; AAB35871.2; -
DR      EMBL; AE008762; AAL20367.1; -
DR      EMBL; AL627271; CAD01922.1; -
DR      Stryene; SG10573; SLTB.
DR      PROSITE; PS00013; PROKAR.LIPOPROTEIN; 1.
KW      Outer membrane; Lipoprotein; Signal; Complete proteome.
FT      SIGNAL
FT      CHAIN
FT      LIPID
FT      LIPID
SQ      SEQUENCE 155 AA; 15548 MW; 82PDDCDBABD55A7 CRC64;
      12.7%; Score 105.5; DB 1; Length 155;
      Best Local Similarity 24.3%; Pred. No. 0.018;
      Matches 34; Conservative 20; Mismatches 53; Indels 33; Gaps 3;
QY      1 MRGLGSSL-----LISFLYGCANFSGEVAATVAGVA 41
DB      15 LACCVNNDLSDDVYTSEAKOVNTYGTIVAVRVYVYLGDDSDNVTAIGAVALGFL 74
QY      42 GQLEFGSGSRVMAIGAVLGLIGSKISGSHDQDKIKLMSLEKACQVTRMNRPP 101
DB      75 GNTIGGSGTRSLATPAAGAVGAVAGVAGOGVSAAMKTVGVEL--EIRK-----DD 120
QY      102 GNSYSEVPRYORYNNOER 121
DB      121 GNTIMVYKQGNTRFSAGQR 140
      -----
RESULT 13
SLTB_ECOLI
ID      SLTB_ECOLI
AC      P55741; P76183; STANDARD; PRT; 155 AA.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Outer membrane lipoprotein sltb precursor.
OS      SLTB OR B1641 OR Z2055 OR ECS2350.
OS      Escherichia coli, and
OC      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
RX      NCBI_TaxID=562, 83334;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9613368; PubMed=8544813;
RA      Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Molienkopf H.-J.,
RA      Goebel W.;
RT      "SLTB, a regulatory protein from Salmonella typhimurium, induces a
RT      haemolytic and pore-forming protein in Escherichia coli.";
RL      Mol. Gen. Genet. 249:474-486(1995).
RP      SEQUENCE FROM N.A.
RX      STRAIN=K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```

```

RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  (3)
RP  SEQUENCE FROM N.A.
RC  STRAIN-K12:
RX  MEDLINE-97251357; Pubmed-9097039;
RA  Alba H., Baba T., Fujita K., Hayashi R., Inada T., Isono K.,
RA  Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA  Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA  Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,
RA  Oshima T., Saito N., Sampa G., Seki Y., Sivasubraman S.,
RA  Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA  Yamamoto Y., Horiuchi T.;
RT  "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT  corresponding to the 28.0-40.1 min region on the linkage map.";
RL  DNA Res. 3:363-377(1996).
RN  (4)
RP  SEQUENCE FROM N.A.
RC  STRAIN-O157:H7 / ATCC 700927;
RX  MEDLINE-21074935; Pubmed-11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA  Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamasis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533(2001).
RN  (5)
RP  SEQUENCE FROM N.A.
RC  STRAIN-O157:H7 / RMD 0509952;
RX  MEDLINE-21156231; Pubmed-11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shida T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
RN  (6)
RP  SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
RP  anchor (Potential).
CC  -1- SIMILARITY: TO S. TYPHIMURUM SLXB, H. INFLUENZAE PCP AND
CC  Y. ENTEROCOLITICA PCP.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AE000259; AAC74713.1; -
DR  EMBL: D90807; BAA15402.1; -
DR  EMBL: AE005387; AAG56630.1; -
DR  EMBL: AP002558; BAB35773.1; -
DR  EcoGene: EG13409; slxb.
DR  PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW  Outer membrane; Lipoprotein; signal; Complete proteome.
CC  FT CHAIN 1 17 POTENTIAL.
CC  FT LIPID 18 155 OUTER MEMBRANE LIPOPROTEIN SLXB.
CC  FT CONFLICT 98 98 N-ACYL DIGLYCERIDE.
CC  FT SEQUENCE 155 AA; 15602 MW; 543EB8A4069A5FA3 CRC64;
SQ
Query Match 12.6%; Score 104.5; DB 1; Length 155;
Best Local Similarity 22.8%; Pred. No. 0.021;
Matches 33; Conservative 21; Mismatches 42; Indels 49; Gaps 4;
QY 12 TTSVFLVCAQNM-----FSROE-----VGAATGAV 36
DB 10 MVLGLVGCNVNDTLSDGVYATSEAKOVQNVSGTIVNRPVQIQGGDDSNVIGALICGAV 69

```

```

QY 37 VGVAVGQLFGKSGRVSMAIGAVLIGLSKIGSQMDQDKIKLNQSLKRVKAGQYTRW 96
DB 70 LGGFLGNTVGGGTGRSLATAGAVAGVAGGCVOSAMNKTOGVEL--EIRK----- 118
QY 97 RNPDTGNSYSVEPRTYQRYNKOER 121
DB 119 ---DDGNTIIVVQKQGNTRFSPQGR 140

RESULT 14
PCP_HAEIN STANDARD; PRT; 155 AA.
ID PCP_HAEIN
AC P10325;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane lipoprotein PCP precursor (15 kDa lipoprotein) (PAL
DE cross-reacting lipoprotein).
GN PCP OR LPP OR H1579.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88115138; Pubmed-2828309;
RA Deich R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
RT outer membrane lipoprotein and an antigenically related 15,000-dalton
RT protein from Haemophilus influenzae.";
RL J. Bacteriol. 170:489-498(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE-95350630; Pubmed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerkavag A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterbach T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURUM SLXB AND TO
CC Y. ENTEROCOLITICA PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: M18877; AAA24938.1; -
DR  EMBL: U32832; AAC23228.1; -
DR  PIR: B28543; B28543.
DR  TIGR: H11579; -
KW  Outer membrane; Lipoprotein; signal; Complete proteome.
CC  FT CHAIN 1 18
CC  FT LIPID 19 155 OUTER MEMBRANE LIPOPROTEIN PCP.
CC  FT CONFLICT 135 143 CSLVAERVF-->VAGARVRI (IN REF. 1).
CC  FT SEQUENCE 155 AA; 15425 MW; D7880327FCF0C985 CRC64;
SQ

```

Query Match 12.1%; Score 101; DB 1; Length 155;  
 Best Local Similarity 39.6%; Pred. No. 0.043;  
 Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 29 VGATGAVGVAGVAGLFGKSGSRVSMATGAVLGLISKIGSMDQDKITL 81  
 Db 62 VGTGGALGTGAGTGGGCGATAAVCAIGALGATGSKIEKMSVNGAEL 114

RESULT 15  
 YCFJ\_ECOLI STANDARD; PRT; 179 AA.  
 ID YCFJ\_ECOLI  
 AC P37951; Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ycfJ.  
 GN YCFJ OR B110.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horichi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [3]  
 RP SEQUENCE OF 1-63 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=81236546; PubMed=6265208;  
 RA Young J.G., Rogers B.L., Campbell H.D., Jaworski A., Shaw D.C.;  
 RT "Nucleotide sequence coding for the respiratory NADH dehydrogenase of  
 Escherichia coli. UUG initiation codon.";  
 RL Eur. J. Biochem. 116:165-170(1981).  
 RN [4]  
 RP IDENTIFICATION.  
 RX MEDLINE=95075659; PubMed=7984428;  
 RA Borodovsky M., Rudd K.E., Koonin E.V.;  
 RT "Intrinsic and extrinsic approaches for detecting genes in a  
 bacterial genome.";  
 RL Nucleic Acids Res. 22:4756-4767(1994).  
 CC -1- SIMILARITY: TO RICKETTSIA 17 KDA SURFACE ANTIGEN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AE000211; AAC74194.1;  
 CC EMBL: D90746; BAA35925.1;  
 CC EMBL: V00306; -; NOT\_ANNOTATED\_CDS.

DR EcoGene; EGI2444; ycfJ.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 5 POTENTIAL.  
 SQ SEQUENCE 179 AA; 18920 MW; BA5EBDB56D45609 CRC64;

Query Match 12.0%; Score 100; DB 1; Length 179;  
 Best Local Similarity 26.3%; Pred. No. 0.061;  
 Matches 36; Conservative 17; Mismatches 48; Indels 36; Gaps 4;

QY 30 GAATGAVGVAGVAGLFGKSGSRVSMATGAVLGLISKIGSMDQDKITLQSLK 89  
 Db 73 GSVLAAGVAGVAGVAGLFGKSGSRVSMATGAVLGLISKIGSMDQDKITLQSLK 120  
 QY 90 AGQVTRMRNPDTGMSYSVEPVRYRYRNKORROQYCRFQKAMIAQKQEL 144  
 Db 121 -----TYTTTQORCKTYVDKSEKMLGIDVYTK-----IGDQGRIRMDRDPG 162  
 QY 145 TACPDPGRMQVISTEK 161  
 Db 163 TQIPL-DSNGQLINNK 178

Search completed: October 27, 2002, 10:56:30  
 Job time : 6.83938 secs

Rickettsial common antigen precursor - Rickettsia conorii

Query Match	34.38;	Score 285;	DB 2;	Length 159;
Best Local Similarity	37.98;			
Matches	61;	Pred. No. 3.1e-18;		

RESULT 3  
A33971

Query Match	34.38;	Score 285;	DB 2;	Length 159
Best Local Similarity	37.98;	Pred. No. 3.1e-18;		
Matches	61;	Conservative	31;	Mismatch

RESULT 4  
G97860

Query Match	34.3%	Score 285;	DB 2;	Length 159;
Best Local Similarity	37.9%	Pred. No. 3,1e-16;		
Matches	61;	Conservative	31;	Mismatches 52;

RESULT 5  
C33971

Query Match	33.28;	Score	276.5;	DB 2;	Length	159;
Best Local Similarity	39.68;	Pred. No.	1.8e-17;			
Matches	55;	Conservative	37;			

RESULT 6  
A25972





C:Genetics:  
A:Gene: PA1053  
C:Superfamily: PAL cross-reacting lipoprotein

Query Match  
Best Local Similarity 13.4%; Score 111.5; DB 2; Length 154;  
Matches 35; Conservative 21; Mismatches 35; Indels 45; Gaps 4;

QY 5 LQSSLLIIS-----VFLVGC-----AQNFSRQE-----  
DB 1 MKKSAIIIVASFTMALALGCGSSLTGDTYSREEARVQVTRMGITQALRPVKEIGTKIP 28  
QY 29 VQAATGAVGVAGVAGQLFGKSGRVSMAIGAVLGLIGSKIGOSMDQDDIKLNOSLEKVK 60  
DB 61 IGSYAGAGVGVAGVAGVAGGKSGSYVAITGAVAGLLGMAITEGILTRTQGEITVREDG 81  
QY 82 --NQSLEKXKAGQVTR 95  
DB 121 STRAYVQVVDGQDIFR 136

# RESULT 11

AE0644

Probable secreted protein STY1252 [Imported] - *Salmonella enterica* subsp. *enterica* serov. *C*:Species: *Salmonella enterica* subsp. *enterica* serovar *Typhimurium*  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 14-Dec-2001  
R:Accession: AE0644  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Moule, S.; Cronin, A.; Davis, P.; Davies, R.M.; Dord, L.; White, N.; Farrar, Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov. *A*:Accession: AE0644  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-179 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08336.1; PID:916502381; GSPDB:GN00176  
C:Gene: STY1252  
C:Superfamily: PAL cross-reacting lipoprotein

Query Match  
Best Local Similarity 13.1%; Score 109; DB 2; Length 179;  
Matches 37; Conservative 16; Mismatches 50; Indels 32; Gaps 4;

QY 30 GAATGAVGVAGVAGQLFGKSGRVSMAIGAVLGLIGSKIGOSMDQDDIKLNOSLEKVK 4;  
DB 73 GSVLGAHVAGGVAGVAGVAGGKSGRVSMAIGAVLGLIGSKIGOSMDQDDIKLNOSLEKVK 89  
QY 90 AGCVTRMRNDPNSVSEPVRTYORNRKORRQOYCCREPO--OKAMTAGOKOETITGTA 120  
DB 121 -----TYYTQORCKTQVYDKSEKMLGYDVYIKIGDQGRKIRMDKP--CTQ 164  
QY 147 CPQPDGRMVOVSTER 161  
DB 165 IPL-DGNQGLVANK 178

# RESULT 12

g87629

Hypothetical protein CC3073 [Imported] - *Caulobacter crescentus*  
C:Species: *Caulobacter crescentus*  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
R:Accession: g87629  
R:Merman, W.C.; Feldlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; Lau, B.; DeBoy, R.T.; Dodson, R.J.; Dutkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon, Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete genome sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: g87629  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-232 <STO>  
A:Cross-references: GB:AE005673; NID:913424723; PIDN:AAK25035.1; GSPDB:GN00148  
C:Gene: CC3073

Query Match  
Best Local Similarity 13.1%; Score 109; DB 2; Length 232;  
Matches 32; Conservative 14; Mismatches 33; Indels 16; Gaps 2;

QY 17 LVGCANFSEKQVAVGAVGVAGVAGQLFGKSGRVSMAIGAVLGLIGSKIGOSMD 73  
DB 81 VVCGKASGKQDEKQVAVGAVGALGAAAGSNLAKNQGCTGTATGAVGAGSLIGCKMKSD 140  
QY 74 DQDQKIKLNQ-----LEKVKAGQVTR 95  
DB 141 MAQEVGGLYKSGGFRVAQTQVAPLVKIEKKVTR 175

# RESULT 13

B82837

Conserved hypothetical protein XF0178 [Imported] - *Xylella fastidiosa* (strain 95sc)  
C:Species: *Xylella fastidiosa*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
R:Accession: B82837  
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Note: for a complete list of authors see reference number A59328 below.  
A:Accession: B82837  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-257 <SIM>

A:Cross-references: GB:AE003872; GB:AE003849; NID:9104975; PIDN:AAF82991.1; GSPDB:GN

A:Experimental source: strain 95sc  
R:Simmons, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
as-Neto, E.; Docena, C.; El-Deiry, H.; Facincani, A.P.; Ferreira, A.J.S.

U.D.; Junqueira, M.L.; Kempner, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
Chado, M.A.; Madeira, A.M.B.N.; Medeira, H.M.F.; Martins, C.L.; Marques, M.V.; Martins  
Rodrigues, V.; Rosa, A.J.; de M.; de Oliveira, M.A.; de Oliveira, R.C.; Palmieri, C.

M.; Teubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:

A:Gene: XF0178  
Query Match  
Best Local Similarity 13.1%; Score 109; DB 2; Length 257;  
Matches 19; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 30 GAATGAVGVAGVAGQLFGKSGRVSMAIGAVLGLIGSKIGOSMDQDDIKLNOSLEKVK 69  
DB 105 GTAIGALIGLIGVGNFGHNGKRAKLTAAAGVAGGFIGNEV 144

# RESULT 14

AD2696

11pA protein [Imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
R:Accession: AD2696  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayvin, T.; Levy, R.; Li, M.; MCCI  
; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AD2696

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-142 <KUR>

A:Cross-references: GB:AE008668; PIDN:AA14986.1; PID:g17739358; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: lipA

A:Map position: circular chromosome

Query Match 12.9%; Score 107.5; DB 2; Length 142;

Best Local Similarity 27.2%; Pred. No. 0.017;

Matches 41; Conservative 20; Mismatches 61; Indels 29; Gaps 9;

QY 9 SLITISVFLVGAQNFNRQEVGAATGAVVGVAGOLFGRSGRSMAGAVLGGIGSK 68

DB 12 SLICVSR-LSAC-----TTTGTTPAG--GSLFGR--SAQSTPFLANLOGGIVG-K 56

QY 69 IGQSMDOODKIKL-----NQSLEKVKAGQVTRWRNPDT-GNSYSVEPVRYQRYNKOERRQ 123

DB 57 SGVELDRGDOTKALEAEYKALETAIPVITWGGDDVKGVANAP---YQVGN----- 107

QY 124 QYCRFPOQKAMINGQKQEIYGTACPOPPGRW 154

DB 108 QNCRQYSHLTVDGRDTRVRGAACRNDGSM 138

# RESULT 15

AE0289

Probable lipoprotein slyB [imported] - Yersinia pestis (strain C092)

C:Species: Yersinia pestis

C:Date: 02-NOV-2001 #sequence\_revision 02-NOV-2001 #text\_change 09-NOV-2001

C:Accession: AF0289

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, L.

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MID:21470413; PMID:11586360

A:Accession: AF0289

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC91178.1; PID:g15980369; GSPDB:GN00175

C:Genetics:

A:Gene: slyB

C:Superfamily: PAL cross-reacting lipoprotein

Query Match 12.9%; Score 107.5; DB 2; Length 155;

Best Local Similarity 24.0%; Pred. No. 0.019;

Matches 36; Conservative 19; Mismatches 44; Indels 51; Gaps 4;

QY 9 SLITISVFLVGAQNFNRQEVGAATGAVVGVAGOLFGRSGRSMAGAVLGGIGSK 68

DB 7 AVAIAVVTLTGCAANNNTLSGDVFTASQAKOVTVSYGTLSVPRVTIOGDDNNVGAIG 66

QY 34 GAVVGGVAGOLFGRSGRSMAGAVLGGIGSKIGQSMDOODKIKL----- 81

DB 67 GAVLGGFLGNAVGGTGRSLATAAGAVAGGIVAGGALNRDGVQLERKDDGQITLV 126

QY 82 --NQSLEKVKAGQVTRWRNPDTGNSYSVEP 109

DB 127 VQKQPTQPSVQGRVMLAN--SGSTITVSP 154

Search completed: October 27, 2002, 10:58:40

Job time: 11.5665 secs

**THIS PAGE BLANK (USPTO)**



Db 491 QGAGAAAAAAGAGGEGIRGAGAGGCGGCGGAGGAGAAAAAGAGAGG 550  
QY 64 LIGSKTQ 71  
Db 551 LGGGAGG 558

## RESULT 2

US-09-247-806-1  
; Sequence 1, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; FILE REFERENCE: 6388-0365-0  
; CURRENT APPLICATION NUMBER: US/09/247,806  
; EARLIER FILING DATE: 1999-02-11  
; EARLIER FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-247-806-1

Query Match 9.4%; Score 78.5; DB 4; Length 651;  
Best Local Similarity 35.3%; Pred. No. 1.7;  
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

QY 27 QEVGATGAVG---GVAQ-----LRKGGGRVSMAGSAVLGG 63  
Db 491 QGAGAAAAAAGAGGEGIRGAGAGGCGGCGGAGGAGAAAAAGAGAGG 550  
QY 64 LIGSKTQ 71  
Db 551 LGGGAGG 558

## RESULT 3

US-08-425-069-2  
; Sequence 2, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5728810th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-425-069-2

Query Match 9.4%; Score 78.5; DB 1; Length 718;  
Best Local Similarity 35.3%; Pred. No. 1.9;  
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

QY 27 QEVGATGAVG---GVAQ-----LRKGGGRVSMAGSAVLGG 63  
Db 491 QGAGAAAAAAGAGGEGIRGAGAGGCGGCGGAGGAGAAAAAGAGAGG 550  
QY 64 LIGSKTQ 71  
Db 551 LGGGAGG 558

## RESULT 4

US-08-317-844B-2  
; Sequence 2, Application US/08317844B  
; Patent No. 5989894  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5989894th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-105P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-317-844B-2

Query Match

9.4%; Score 78.5; DB 2; Length 718;

Best Local Similarity 35.3%; Pred. No. 1.9;  
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

OY 27 QEVGATGATGAVG---GVAGQ-----LFGKSGSRVSMATGAVLGG 63  
Db 491 QGAGAAAAAAGAGGEGIRGGAGGGYGGIGSGSGRGLGGGAGAAAAAGAGGCGG 550  
OY 64 LIGSKTGO 71  
Db 551 LGGGAGG 558

## RESULT 5

US-09-034-177-3  
; Sequence 3, Application US/09034177  
; Patent No. 6127146  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; City: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/034,177  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0486 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: GI 1174414  
; US-09-034-177-3

Query Match 9.4%; Score 78.5; DB 3; Length 747;  
Best Local Similarity 35.3%; Pred. No. 2;  
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

OY 27 QEVGATGATGAVG---GVAGQ-----LFGKSGSRVSMATGAVLGG 63  
Db 491 QGAGAAAAAAGAGGEGIRGGAGGGYGGIGSGSGRGLGGGAGAAAAAGAGGCGG 550  
OY 64 LIGSKTGO 71  
Db 551 LGGGAGG 558

RESULT 6  
US-09-553-498-8  
; Sequence 8, Application US/09553498  
; Patent No. 6309861  
; GENERAL INFORMATION:  
; APPLICANT: Ambrosius, Dorothee  
; APPLICANT: Rudolph, Rainer  
; APPLICANT: Schaeffner, Joerg  
; APPLICANT: Schwarz, Elisabeth  
; TITLE OF INVENTION: Process for the production of naturally folded and secreted pr  
; FILE REFERENCE: Case 20379  
; CURRENT APPLICATION NUMBER: US/09/553,498  
; CURRENT FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: EP99107412.1  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 10  
; SEQ ID NO 8  
; LENGTH: 255  
; TYPE: PR  
; ORGANISM: E. coli  
; US-09-553-498-8

Query Match 9.3%; Score 77.5; DB 4; Length 255;  
Best Local Similarity 20.9%; Pred. No. 0.62;  
Matches 37; Conservative 28; Mismatches 81; Indels 31; Gaps 7;

OY 5 LQSSLIIVELVGCANFSRQEVGATGAV---GVAGQLFGKSGSRVSMATGAV 60  
Db 83 LQMTSLRSEDTAMYYCARDGAY-WGQGTIVYSSGGGSGGGSGLDIETQSPAI 141  
OY 61 LGLIGSKIGSGMDQDKIK-LN-----QSLKRYKAGVTRMRNDTGSY 105  
Db 142 MSASPGKVTWTCSSASSVRMMWFQKSGTSPKRWIYDPSKSSGVPARSGSGSTSY 201  
OY 106 SVEPRTYRKNQERYQYCRFQKRAMI--AGQKQEIYGTACPPDGRMYISTE 160  
Db 202 SL---TTSMEADATYTCQQWNSNPLFGAGTKLELRAAAEQ-----KLISEE 249

## RESULT 7

US-09-067-351-2  
; Sequence 2, Application US/09067351  
; Patent No. 5994081  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Baughn, Mariah  
; TITLE OF INVENTION: HUMAN KERATINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/067,351  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CERRONE, MICHAEL C.  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0511 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02  
CLONE: 2029060  
US-09-067-351-2

Query Match  
Best Local Similarity 9.3%; Score 77; DB 2; Length 551;  
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 7 GSSLIITSVFLVCAQNFSEVGAATGAVGVAGOLFSGKSGRVSMATGAVLGGLIG 66  
DB 54 GASFGSRSLYNLGAKRVSILNCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110  
QY 67 SK-----IGSMDDQDKIKLNQSLSEKVKAGQVTRMRNPDTGNSYSEVPRTY 113  
DB 111 PSFVPCPGGIGQEVTVNOSLLTPHLQIDPTIQVRAERERQIKTLNKKFTSFIDKVRFL 170  
QY 114 QRYNK 118  
DB 171 EQONK 175

## RESULT 8

US-09-360-490-2  
Sequence 2, Application US/09360490  
Patent No. 6221843

GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Marian  
TITLE OF INVENTION: HUMAN KERATINS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/360,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/067,351  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0511 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 551 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: KERANOT02

CLONE: 2029060  
US-09-360-490-2

Query Match  
Best Local Similarity 9.3%; Score 77; DB 4; Length 551;  
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 7 GSSLIITSVFLVCAQNFSEVGAATGAVGVAGOLFSGKSGRVSMATGAVLGGLIG 66  
DB 54 GASFGSRSLYNLGAKRVSILNCGSSCRSGFGGRASNGFVNSG---FGYGGVGGGFGSG 110  
QY 67 SK-----IGSMDDQDKIKLNQSLSEKVKAGQVTRMRNPDTGNSYSEVPRTY 113  
DB 111 PSFVPCPGGIGQEVTVNOSLLTPHLQIDPTIQVRAERERQIKTLNKKFTSFIDKVRFL 170  
QY 114 QRYNK 118  
DB 171 EQONK 175

## RESULT 9

US-08-374-077C-2  
Sequence 2, Application US/08374077C  
Patent No. 6027912

GENERAL INFORMATION:  
APPLICANT: Hall, Linda M.  
APPLICANT: Ren, Dejian  
APPLICANT: Zheng, Wei  
APPLICANT: Dubaid, Manuel Marcel Paul  
TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/374,077C  
FILING DATE: 19-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm M.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 022650-264  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2516 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-374-077C-2

Query Match  
Best Local Similarity 9.2%; Score 76.5; DB 3; Length 2516;  
Matches 32; Conservative 16; Mismatches 66; Indels 13; Gaps 4;

QY 29 VGATGAVGVAGOLFSGKSGRVSMATGAVLGGLIGSGTIGS-----MDQDKIKLN 82  
DB 2389 IGSSNGSIFGSGAGLGAGSGGVG-GLGSSSIRNAFGSGSGPSSLRQHPPTSGTIN 2447  
QY 83 -----OSLEKVKAGQVTRMRNPDTGNSYSEVPRTYQRNRQERQOYCRFGQKAMIA 136



Db 2448 SPPIDNRLRLRVATVTTNNNNKSOVSQNNSSLNVRANANSQNMSPGQPVQOQSPLR 2507  
QY 137 GOKOEIYGT 145  
Db 2508 GQGNQYSS 2516

## RESULT 10

US-08-895-590-2  
; Sequence 2, Application US/08895590  
; Patent No. 6207410  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubald, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,590  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374,888  
; FILING DATE: 19-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm M.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 022650-263  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;  
Best Local Similarity 24.8%; Pred. No. 17;

Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

QY 29 VGAATGAVGVAGQLFGKSGRYSAIGG--AVLGLIGSKIGS---MDQODKIKLN 82  
Db 2389 IGSSNGSIFGSGAGLGGAGSGVG-GLGSSSIRNMFSGSGSPSLSPQHQPYSGLN 2447  
QY 83 -----OSLEVKVAGVTRMRNPDTGNSYSVEPRITQRYNKKQRRQYCFEFOOKAMIA 136  
Db 2448 SPPIDNRLRLRVATVTTNNNNKSOVSQNNSSLNVRANANSQNMSPGQPVQOQSPLR 2507  
QY 137 GOKOEIYGT 145  
Db 2508 GQGNQYSS 2516

RESULT 11  
US-07-803-633A-13  
; Sequence 13, Application US/07803633A  
; Patent No. 5369025

## GENERAL INFORMATION:

; APPLICANT: NAZERIAN, Keyvan  
; APPLICANT: LEE, Lucy F.  
; APPLICANT: YANAGIDA, No. 5369025oru  
; APPLICANT: OGAMA, Ryohel  
; APPLICANT: LI, Yi  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VACCINE FOR  
; PROTECTION AGAINST MAREK'S DISEASE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 301 No. 5369025th Washington street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/803,633A  
; FILING DATE: 19911210  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1644-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 865 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-803-633A-13

Query Match 9.1%; Score 76; DB 1; Length 865;

Best Local Similarity 25.9%; Pred. No. 4.7; Mismatches 46; Indels 30; Gaps 8;

Matches 37; Conservative 30; Mismatches 46; Indels 30; Gaps 8;

QY 16 FLVCAQNFSSRO-EVCAATGAVGVAGQLFGKSGRYSAIGG--AVLGLIGSKIGS---MDQODKIKLN 82  
Db 688 FMNLAELEFNGMGVGAIGVAVGAGAIYSTISG-VSAFMSNPFGLAIGLIINGLV 746  
QY 66 GSKIGSDQODKIKLN-----OSLEVKVAGVTRMRNPDTGNSYSVEPRITQRYNKK 118  
Db 747 AAFI--AVRYVNRKLKSNPKALYPMTEVLKA-QATRELGHEESDD-----LETSI 795  
QY 119 QERROQYCFEFOOKAMIA 140  
Db 796 DERKLEAREMIKYMALVSABER 818

## RESULT 12

US-08-220-151-6  
; Sequence 6, Application US/08220151  
; Patent No. 5529780  
; GENERAL INFORMATION:  
; APPLICANT: Paolelli, Enzo  
; APPLICANT: Limbach, Keith J.  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
; CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY

```

      COUNTRY: USA
      ZIP: 10036
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/220,151
      FILING DATE: 30-MAR-1994
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
      NAME: Frommer, William S.
      REGISTRATION NUMBER: 25,506
      REFERENCE/DOCKET NUMBER: 454310-2540
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (212) 840-3333
      TELEFAX: (212) 840-0712
      INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 913 amino acids
      TYPE: amino acid
      STRANDEDNESS: Single
      TOPOLOGY: linear
      MOLECULE TYPE: peptide
      FRAGMENT TYPE: N-terminal
      US-08-220-151-6

Query Match
Best Local Similarity 23.68; Score 74.5; DB 1; Length 913;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

OY 15 VFLVCAQNF-SROEVGATGAVGVAAGOLFEGKSGRVS-----MAIGAVLGGLI 65
DB 753 VLLRGIANFPGGLDVGAAVGVKVLGATGAVISAVGAVSFLSNPGALAIGLVLGLV 812
OY 66 GSKI-----GOSMDQDKIKLNOSLEKVK-----A 90
DB 813 AAFLAYHISRLRNPMKALYPVTTKLEDEVDGDEAKLDQARMIRMSIVSALE 872
OY 91 GQVTRRNPDGTGNSYSEPV-----RTYORYNKOE 120
DB 873 QQEHKARKKNSGALLASRVGAMATRRRHYQRLSESD 909

RESULT 13
US-08-413-118-6
; Sequence 6, Application US/08413118
; Patent No. 5688920
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: LIMBACH, KEITH J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE, 25TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,118
; FILING DATE: 29-MAR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

      APPLICATION NUMBER: US 08/220,151
      FILING DATE: 30-MAR-1994
      ATTORNEY/AGENT INFORMATION:
      NAME: FROMMER, WILLIAM S.
      REGISTRATION NUMBER: 25,506
      REFERENCE/DOCKET NUMBER: 454310-2670
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (212) 840-3333
      TELEFAX: (212) 840-0712
      INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 913 amino acids
      TYPE: amino acid
      STRANDEDNESS: Single
      TOPOLOGY: linear
      MOLECULE TYPE: peptide
      FRAGMENT TYPE: N-terminal
      US-08-413-118-6

Query Match
Best Local Similarity 23.68; Score 74.5; DB 1; Length 913;
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

OY 15 VFLVCAQNF-SROEVGATGAVGVAAGOLFEGKSGRVS-----MAIGAVLGGLI 65
DB 753 VLLRGIANFPGGLDVGAAVGVKVLGATGAVISAVGAVSFLSNPGALAIGLVLGLV 812
OY 66 GSKI-----GOSMDQDKIKLNOSLEKVK-----A 90
DB 813 AAFLAYHISRLRNPMKALYPVTTKLEDEVDGDEAKLDQARMIRMSIVSALE 872
OY 91 GQVTRRNPDGTGNSYSEPV-----RTYORYNKOE 120
DB 873 QQEHKARKKNSGALLASRVGAMATRRRHYQRLSESD 909

RESULT 14
US-08-473-446-6
; Sequence 6, Application US/08473446
; Patent No. 6017542
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: LIMBACH, KEITH J.
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE, 25TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/413,118
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FROMMER, WILLIAM S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2670
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 6:

```

SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-473-446-6

Query Match 9.0%; Score 74.5; DB 3; Length 913;  
Best Local Similarity 23.6%; Pred. No. 7.4;  
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCANF-SROEVGATGAVGVAGOLFSGSRVS-----MAIGAVLGGLI 65  
| | | | | : | | | | | : | | : | | : | | : | | :  
DB 753 VLLRGIANFPGAGDVGAAVGVKVLGATGAVISAGVWSFLSNPFGALAIGLLVLAGLV 812  
QY 66 GSKI-----GQSMDOODKIKLNSLEKYK-----A 90  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 813 AAFLAYRHISRLRRNPKALYPVTTKTKLEDGVDEGVDEAKLDQARDMIRYMSIVSALE 872  
QY 91 GQYTRMRNPDTGNSYVEPV-----RTYQRYNKQE 120  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 873 QOEHRKARKKNSGALLASRVGAMATRRRRHYORLESED 909

## RESULT 15

US-09-232-468A-2  
Sequence 2, Application US/09232468A  
Patent No. 6207165  
GENERAL INFORMATION:  
APPLICANT: AUDONNET et al.  
TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE  
FILE REFERENCE: 454313-2230  
CURRENT APPLICATION NUMBER: US/09/232.468A  
CURRENT FILING DATE: 1999-01-05  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 913  
TYPE: PRT  
ORGANISM: Pseudorabies virus  
US-09-232-468A-2

Query Match 9.0%; Score 74.5; DB 4; Length 913;  
Best Local Similarity 23.6%; Pred. No. 7.4;  
Matches 37; Conservative 18; Mismatches 51; Indels 51; Caps 5;

QY 15 VFLVGCANF-SROEVGATGAVGVAGOLFSGSRVS-----MAIGAVLGGLI 65  
| | | | | : | | | | | : | | : | | : | | : | | :  
DB 753 VLLRGIANFPGAGDVGAAVGVKVLGATGAVISAGVWSFLSNPFGALAIGLLVLAGLV 812  
QY 66 GSKI-----GQSMDOODKIKLNSLEKYK-----A 90  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 813 AAFLAYRHISRLRRNPKALYPVTTKTKLEDGVDEGVDEAKLDQARDMIRYMSIVSALE 872  
QY 91 GQYTRMRNPDTGNSYVEPV-----RTYQRYNKQE 120  
| : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 873 QOEHRKARKKNSGALLASRVGAMATRRRRHYORLESED 909

Search completed: October 27, 2002, 11:00:28  
Job time : 9.34197 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 ; Search time 21.6891 Seconds  
(without alignments)  
824.509 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832  
Sequence: 1 MRCGLQGSSLIITISVFLVNC.....ITGTACPPQDGRWQVISTEK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	832	100.0	161	22	Optimised Ospa pro
2	832	100.0	256	22	CI7E2 Ospa constu
3	815	98.0	162	22	Piscirickettsia sa
4	815	98.0	162	22	Ospa antigen amino
5	112	13.5	20	22	Ospa B-cell epitop
6	112	13.3	224	22	Moraxella catarrha
7	102.5	12.3	223	20	Porphyromonas ging
8	102.5	12.3	230	20	Porphyromonas ging
9	101	12.1	154	11	PBOMP-2 gene prod.
10	95	11.4	309	22	Novel human diagno
11	88	10.6	528	22	Spider recombinant

12	87.5	10.5	147	22	ABB69847
13	86.5	10.4	2309	22	ABB66232
14	84	10.1	1251	22	ABB61254
15	83	10.0	116	19	AAV11028
16	83	10.0	423	22	ABG30695
17	82	9.9	2017	22	ABG06301
18	81	9.7	666	22	ABBS8019
19	80.5	9.7	542	22	ABB65790
20	80.5	9.7	542	22	ABB65791
21	80.5	9.7	542	22	ABB70501
22	79.5	9.6	618	21	AAV56803
23	78.5	9.4	651	20	AAV40097
24	78.5	9.4	718	12	AAV14308
25	78.5	9.4	718	19	AAV53346
26	78.5	9.4	718	21	AAV59070
27	77.5	9.3	102	22	AAV41943
28	77.5	9.3	255	21	AAV11398
29	77.5	9.3	255	22	AAV74199
30	77.5	9.3	255	22	AAV70769
31	77.5	9.3	255	22	AAV72020
32	77.5	9.3	285	20	AAV41688
33	77.5	9.3	285	21	AAV44244
34	77.5	9.3	285	22	AAV29025
35	77.5	9.3	285	22	AAV39011
36	77.5	9.3	302	22	AAV40157
37	77.5	9.3	354	22	AAV19445
38	77.5	9.3	514	22	AAV36520
39	77	9.3	223	21	AAV57121
40	77	9.3	285	21	AAV87280
41	77	9.3	285	21	AAV73440
42	77	9.3	285	22	AAV53288
43	77	9.3	251	21	AAV52398
44	77	9.3	259	21	AAV75098
45	76.5	9.2	2516	17	AAV01875

#### ALIGNMENTS

RESULT 1	
AAV81127	
ID AAV81127 standard; Protein; 161 AA.	
XX	
AC AAV81127;	
XX	
DT 11-JUL-2001 (first entry)	
XX	
DE Optimised Ospa protein 17E2 amino acid sequence.	
XX	
KW Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;	
KW vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;	
KW SRS.	
XX	
OS Piscirickettsia salmonis.	
OS Synthetic.	
XX	
FH Key	Location/Qualifiers
FT Region	109..128
FT	/label= B_cell_epitope
XX	
PN CA2281913-A1.	
XX	
PD 17-MAR-2001.	
XX	
PF 17-SEP-1999; 99CA-2281913.	
XX	
PR 17-SEP-1999; 99CA-2281913.	
XX	
PA (KAYW/) KAY W W.	
PA (BURI/) BURIAN J.	
PA (KUZV/) KUZV M A.	
XX	
PI Kay WW, Burian J, Kuzv MA;	

Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
H. pylori ORF 01c  
A fusion of anti-C  
Novel human diagno  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Human prostate can  
Human prostate can  
Spider silk protei  
N. clavipes draglin  
Nephila clavipes s  
N. clavipes spider  
Human polypeptide  
E. coli expression  
PelB-scFvOxazol  
Expression plasmid  
E. carotovora PelB  
Human PRO284 prote  
Human PRO284 (UNO2  
Human PRO polypept  
Human polypeptide  
Human polypeptide  
Human diagnostic a  
Pseudomonas aerugi  
Human prostate can  
Human signal pepti  
Human secreted pro  
Human protein HPI0  
Human keratin KERK  
Neisseria meningit  
Neuronal Invertebr

```

XX DR WPI; 2001-316844/34.
XX N-PSDB; AAF86247.
XX
XX PT Method for protecting polkilothermic fish against salmonid rickettsial
XX PT septicæmia and other rickettsial diseases comprises administering a
XX PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX PS
XX PS Disclosure: Fig 5; 35pp; English.
XX
XX CC This invention relates to a method for the protection against infection
XX CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia
XX CC P. salmonis. The method comprises administering an immunogenic amount of a
XX CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
XX CC Ospa in the form of a vaccine. The method is used for protecting animals,
XX CC particularly polkilothermic fish, against the bacterial pathogen
XX CC P. salmonis. The method is also useful for protecting against salmonid
XX CC rickettsial septicæmia (SRS) and other rickettsial diseases. The present
XX CC sequence represents optimised P. salmonis Ospa protein 17E2. The DNA
XX CC encoding Ospa 17E2 (AAF86247) has been optimised for expression in
XX CC Escherichia coli. An Ospa protein with an N-terminal fusion partner is
XX CC used in a vaccine to create an anti-Ospa antibody response.
XX
XX SQ Sequence 161 AA;
XX
XX Query Match 100.0%; Score 832; DB 22; Length 161;
XX Best Local Similarity 100.0%; Pred. No. 3e-81;
XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MRGCLGGSSLLIIISVFLVGCAGNFSROEVGAATGAVGVAGOLFPGKSGRVSMAIGAV 60
XX DB 1 MGCCLOGSSLLIIISVFLVGCAGNFSROEVGAATGAVGVAGOLFPGKSGRVSMAIGAV 60
XX QY 61 LGGLIGSKIGSMDOQDKIKLNSLEKVKAGGYTRMRNPDGNSYSVEPVRTYQRYNQE 120
XX DB 61 LGGLIGSKIGSMDOQDKIKLNSLEKVKAGGYTRMRNPDGNSYSVEPVRTYQRYNQE 120
XX QY 121 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWQYISTEK 161
XX DB 121 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWQYISTEK 161
XX
XX RESULT 2
XX ID AAB81128 standard; Protein: 256 AA.
XX AC AAB81128;
XX
XX DT 11-JUL-2001 (first entry)
XX DE C17E2 Ospa construct with N-terminal fusion partner.
XX
XX KW Polkilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX KW vaccine; Ospa; salmonid rickettsial septicæmia; rickettsial disease;
XX KW SRS; 17E2; fusion construct.
XX
XX OS Piscirickettsia salmonis.
XX OS Synthetic.
XX
XX FT Key
XX FT Region 1..95
XX FT Location/Qualifiers
XX FT Region 96..256
XX FT /label= C17E2_Ospa
XX FT /note= "Product of Ospa gene optimised for expression in
XX FT Escherichia coli."
XX
XX CA2281913-A1.
XX
XX PD 17-MAR-2001.
XX
XX PF 17-SEP-1999; 99CA-2281913.
XX

```

```

PR 17-SEP-1999; 99CA-2281913.
XX
XX PA (KAYW/) KAY W W.
XX PA (BURJ/) BURIAN J.
XX PA (KUZU/) KUZUK M A.
XX
XX PI Kay WW, Burian J, Kuzuk MA;
XX
XX DR WPI; 2001-316844/34.
XX N-PSDB; AAF86248.
XX
XX PT Method for protecting polkilothermic fish against salmonid rickettsial
XX PT septicæmia and other rickettsial diseases comprises administering a
XX PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX PS
XX PS Example 4; Fig 5; 35pp; English.
XX
XX CC This invention relates to a method for the protection against infection
XX CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia
XX CC P. salmonis. The method comprises administering an immunogenic amount of a
XX CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
XX CC Ospa in the form of a vaccine. The method is used for protecting animals,
XX CC particularly polkilothermic fish, against the bacterial pathogen
XX CC P. salmonis. The method is also useful for protecting against salmonid
XX CC rickettsial septicæmia (SRS) and other rickettsial diseases. The present
XX CC sequence represents the amino acid sequence of C17E2, a P. salmonis Ospa
XX CC construct optimised for expression in Escherichia coli, fused to an
XX CC undefined N-terminal fusion partner. The fusion protein is used in a
XX CC vaccine to create an anti-Ospa antibody response.
XX
XX SQ Sequence 256 AA;
XX
XX Query Match 100.0%; Score 832; DB 22; Length 256;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-81;
XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MRGCLGGSSLLIIISVFLVGCAGNFSROEVGAATGAVGVAGOLFPGKSGRVSMAIGAV 60
XX DB 96 MRGCLGGSSLLIIISVFLVGCAGNFSROEVGAATGAVGVAGOLFPGKSGRVSMAIGAV 60
XX QY 61 LGGLIGSKIGSMDOQDKIKLNSLEKVKAGGYTRMRNPDGNSYSVEPVRTYQRYNQE 155
XX DB 156 LGGLIGSKIGSMDOQDKIKLNSLEKVKAGGYTRMRNPDGNSYSVEPVRTYQRYNQE 120
XX QY 121 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWQYISTEK 161
XX DB 216 RROQYCREFOOKAMIAQOKOEIYGTACPOPdGRWQYISTEK 256
XX
XX RESULT 3
XX ID AAG78025 standard; Protein: 162 AA.
XX AC AAG78025;
XX
XX DT 15-JAN-2002 (first entry)
XX DE Piscirickettsia salmonis polypeptide P10.6.
XX
XX KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
XX KW septicæmia; SRS; surface antigen; vaccine; antibacterial; fish;
XX KW ATCC VR-1361.
XX
XX OS Piscirickettsia salmonis.
XX
XX PN WO200168865-A2.
XX
XX PD 20-SEP-2001.
XX
XX PF 12-MAR-2001; 2001MO-GB01055.
XX
XX PR 11-MAR-2000; 2000GB-0005838.
XX PR 01-JUL-2000; 2000GB-0016080.
XX

```

```

PR 01-JUL-2000: 2000GB-0016082.
PR 29-JUL-2000: 2000GB-0018599.
XX
XX (AQUA-) AQUA HEALTH EURO LTD.
XX
XX Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI Burzio L;
XX
XX WPI: 2001-639050/73.
DR N-PSDB; AAF89040.
XX
XX New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on Piscirickettsia salmonis are useful to
PT protect fish against Piscirickettsiosis
XX
XX Claim 6; Fig 5; 25pp; English.
XX
XX The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on Piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.
XX
XX
SQ Sequence 162 AA:

Query Match 98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79; Mismatches 1; Indels 0; Gaps 0;
Matches 158; Conservative 1;

QY 2 RGLGSSLLIIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGSGRVSMAGAVL 61
DB 3 RGLGSSLLIIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGSGRVSMAGAVL 62
QY 62 GGLIGSKIGSQMDQDKIKLNQSLSEKVKAGQVTRMRNPDGNSYSVEPRTYORYNKOER 121
DB 63 GGLIGSKIGSQMDQDKIKLNQSLSEKVKAGQVTRMRNPDGNSYSVEPRTYORYNKOER 122
QY 122 ROOYCREFOOKAMTAGOKOIEYGTACROPDGRMOVISTEK 161
DB 123 ROOYCREFOOKAMTAGOKOIEYGTACROPDGRMOVISTEK 162

RESULT 4
AAB81126
ID AAB81126 standard; Protein; 162 AA.
XX
XX AAB81126;
AC
XX
DT 11-JUL-2001 (first entry)
XX
XX OSPA antigen amino acid sequence.
DE
XX
XX Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
KM SRS.
XX
XX Piscirickettsia salmonis.
OS
XX
XX Key Location/Qualifiers
FT Region 110..129
FT /label= B_cell_epitope
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZV M A.
XX

```

```

XX
XX Kay WM, Burian J, Kuzyk MA;
PI
XX
XX WPI: 2001-316844/34.
DR N-PSDB; AAF86246.
XX
XX Method for protecting polkillothermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the OSPA protein of Piscirickettsia salmonis
XX
XX Example 2; Fig 2b; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkillothermic fish by the bacterial pathogen, Piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
CC OSPA in the form of a vaccine. The method is used for protecting animals,
CC particularly polkillothermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents P. salmonis OSPA protein. An OSPA protein with an
CC N-terminal fusion partner is used in a vaccine to create an anti-OSPA
XX antibody response.
XX
SQ Sequence 162 AA:

Query Match 98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79; Mismatches 1; Indels 0; Gaps 0;
Matches 158; Conservative 1;

QY 2 RGLGSSLLIIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGSGRVSMAGAVL 61
DB 3 RGLGSSLLIIISVFLGCAQNFROEVGAATGAVVGAVGOLFSGSGRVSMAGAVL 62
QY 62 GGLIGSKIGSQMDQDKIKLNQSLSEKVKAGQVTRMRNPDGNSYSVEPRTYORYNKOER 121
DB 63 GGLIGSKIGSQMDQDKIKLNQSLSEKVKAGQVTRMRNPDGNSYSVEPRTYORYNKOER 122
QY 122 ROOYCREFOOKAMTAGOKOIEYGTACROPDGRMOVISTEK 161
DB 123 ROOYCREFOOKAMTAGOKOIEYGTACROPDGRMOVISTEK 162

RESULT 5
AAB81130
ID AAB81130 standard; Peptide; 20 AA.
XX
XX AAB81130;
AC
XX
DT 11-JUL-2001 (first entry)
XX
XX OSPA B-cell epitope peptide #2.
DE
XX
XX Polkillothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
KM SRS; antibody.
XX
XX Piscirickettsia salmonis.
OS
XX
XX Key Location/Qualifiers
FT Region 110..129
FT /label= B_cell_epitope
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZV/) KUZV M A.
XX
XX Kay WM, Burian J, Kuzyk MA;
PI
XX

```







Db

62 VETLGGALGAGIAGTIGGGRGQAIAAVGAIGALAGSKIEKMSQVNGAEL 114

RESULT 10

ABG15906  
ID ABG15906 standard; Protein; 309 AA.  
XX  
AC ABG15906;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15897.  
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR MPI: 2001-639362/73.  
XX  
DR N-PSDB: AAS80093.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
XX biodiversity  
XX

Claim 20; SEQ ID NO 46265; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX and to produce other types of data and products dependent on mutations  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX

Sequence 309 AA;

Query Match 11.4%; Score 95; DB 22; Length 309;  
Best Local Similarity 33.0%; Pred. No. 0.077;  
Matches 32; Conservative 12; Mismatches 39; Indels 14; Gaps 3;

QY 12 ITSVFLVGAQNFNR-----GEVGAATGAVGVAGQLFGKG--SGRVSMATGAVAGL 64  
DB 185 ITSLPAAVAVDQNSKRNKAQAFALIGAVAGVIGHNVSNSGTTAGAVGGAVGAA 244

QY 65 IGSKI-----GOSMDQDKIKLNSLEKVRKAGV 94  
DB 245 AGSMVNDKTIAMEGVSLTYKRGTRVYTSVGKEKFQFT 281

RESULT 11

AAB82611  
ID AAB82611 standard; Protein; 528 AA.  
XX  
AC AAB82611;  
XX  
DT 02-OCT-2001 (first entry)  
XX  
DE Spider recombinant silk protein PETNCDs.  
XX  
DE Spider: orb-weaver; silk protein; PETNCDs; structural protein;  
XX purification; fibre; spinning.  
XX  
OS Nephila clavipes.  
XX

XX Key Location/Qualifiers  
XX FT Misc-difference 417  
XX FT /note= "encoded by TAT"  
XX FT Misc-difference 427  
XX FT /note= "encoded by CGA"  
XX FT Misc-difference 522  
XX FT /note= "encoded by GAG"  
XX  
XX WO200153333-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 01-NOV-2000; 2000WO-US30086.  
XX  
XX 20-JAN-2000; 2000US-0490291.  
XX  
XX (MELI/) MELLO C M.  
XX (ARCI/) ARCIDIAONO S.  
XX (BUTL/) BUTLER M M.  
XX (USSA ) US SEC OF ARMY.  
XX

PI Mello CM, Arcidiacono S, Butler MM;  
XX  
XX MPI: 2001-483136/52.  
XX  
XX N-PSDB: AAB26304.  
XX

PT Recovering structural polypeptides in a biological sample, useful for  
PT purifying and spinning spider silks and other structural proteins, the  
PT comprises treating the sample containing the polypeptides with an acid  
XX  
XX Claim 2; Page 41-42; 49p; English.

XX The present sequence is that of the orb-weaver spider (Nephila  
XX clavipes) recombinant silk protein PETNCDs. The invention  
XX provides methods for purifying and spinning spider silks and other  
XX structural proteins. Organic acids are used to lyse recombinant  
XX cells or other biological samples (such as non-recombinant  
XX derived cells), and enrich the purity and yields of structural  
XX proteins by hydrolysing many of the macromolecules while leaving  
XX the structural proteins intact. In the case of silk proteins, the  
XX resulting lysate is further purified by ion-exchange or affinity  
XX chromatography and processed into an aqueous-based mixture for  
XX into vector pET24 for recombinant expression in Escherichia coli,  
XX containing denaturant (3 M guanidine-HCl) and by affinity  
XX chromatography on nickel-NiA agarose. Products obtained using the  
XX methods of the invention can be used in the construction of many  
XX materials including films, fibres, woven articles, sutures, a  
XX method has the following advantages over prior art: it involves  
XX fewer steps, requires less time and smaller volumes of reagents,

CC results in better recovery of protein at higher purity (70-99%),  
CC is easy to scale up, and the fibres are spun in an environmentally  
CC benign solution reducing hazardous waste accumulation and cost.

XX Sequence 528 AA:

Query Match 10.6%; Score 88; DB 22; Length 528;  
Best Local Similarity 48.9%; Pred. No. 0.87;  
Matches 23; Conservative 3; Mismatches 19; Indels 2; Gaps 1;

OY 27 QEYGAATGAVGVCAGQ--LFGKSGSRVSMATGAVLGLIGSKIQ 71  
DB 320 QGAGAAAAAAGGAGCGGIGCGAGAAAAAAGAGCGGIGCGAGQ 366

RESULT 12

ABB69847  
ID ABB69847 standard; Protein: 147 AA.

XX ABB69847;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 36333.

KW Drosophila: developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PA Venter JC, Adams M, Li PMD, Myers EM;

PI WPI: 2001-656860/75.

DR N-PSDB; ABL13950.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

PS Disclosure: SEQ ID NO 36333; 21pp + Sequence Listing: English.

CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins  
sequences (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 147 AA:

Query Match 10.5%; Score 87.5; DB 22; Length 147;  
Best Local Similarity 27.5%; Pred. No. 0.19;  
Matches 36; Conservative 18; Mismatches 38; Indels 39; Gaps 7;

OY 36 VVGAVAGLFGKSGSRVSMATGAVLGLIGSKIGSQMDQ---DKIKLN 82  
DB 11 VVGAVAGLFGKSGSRVSMATGAVLGLIGSKIGSQMDQ---DKIKLN 82

OY 83 QSLKVKAGQWTRMRNDGNSYSVEPRTYORYNKORROQYCR-EFOOKAMIAQKOE 141  
DB 69 QSLKVKAGQWTRMRNDGNSYSVEPRTYORYNKORROQYCR-EFOOKAMIAQKOE 141

OY 142 YGTACPPPDG 152

DB 107 YGTACPPPDG 116

RESULT 13

ABB66232  
ID ABB66232 standard; Protein: 2309 AA.

XX ABB66232;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 25488.

KW Drosophila: developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PA Venter JC, Adams M, Li PMD, Myers EM;

PI WPI: 2001-656860/75.

DR N-PSDB; ABL10335.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

PS Disclosure: SEQ ID NO 25488; 21pp + Sequence Listing: English.

CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins  
sequences (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2309 AA:

Query Match 10.4%; Score 86.5; DB 22; Length 2309;  
Best Local Similarity 27.4%; Pred. No. 8.5;  
Matches 34; Conservative 18; Mismatches 55; Indels 17; Gaps 5;

OY 22 QNFSRQEVGAATGA---VVGAVAGLFGKSGSRVSMATGAVLGLIGSKIGSQMDQDK 78

DB 2124 QNFSRQEVGAATGA---VVGAVAGLFGKSGSRVSMATGAVLGLIGSKIGSQMDQDK 78

OY 79 IKANQSLKVKAGQWTRMRNDGNSYSVEPRTYORYNKORROQYCR-EFOOKAMIAQ 137

DB 2179 GSVGQSLKVKAGQWTRMRNDGNSYSVEPRTYORYNKORROQYCR-EFOOKAMIAQ 137



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:01 : Search time 18.0743 Seconds  
(without alignments)  
1540.986 Million cell updates/sec

Title: US-09-677-374-4  
Perfect score: 832  
Sequence: 1 MRGCLQGSSLLIITSVFLVGC.....ITGTACPPDPGRMOWISTEK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SP archaea:\*
  - 2: SP bacteria:\*
  - 3: SP fungi:\*
  - 4: SP human:\*
  - 5: SP invertebrate:\*
  - 6: SP mammal:\*
  - 7: SP mhc:\*
  - 8: SP organelle:\*
  - 9: SP phage:\*
  - 10: SP plant:\*
  - 11: SP rodent:\*
  - 12: SP virus:\*
  - 13: SP vertebrate:\*
  - 14: SP unclassified:\*
  - 15: SP\_rvirus:\*
  - 16: SP\_bacteriap:\*
  - 17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	815	98.0	162	2	09F9K8	pisciricket
2	303.5	36.5	148	2	054381	rickettsia
3	283	34.0	159	2	09F9F2	rickettsia
4	258.5	31.1	137	2	052252	rickettsia
5	252.5	30.3	144	2	09K2N6	male-Killin
6	251.5	30.2	137	2	031065	rickettsia
7	251.5	30.2	144	2	09K4W8	male-Killin
8	244	29.3	154	2	053154	rickettsia
9	239	28.7	151	2	09F9Q9	rickettsia
10	236.5	28.4	131	2	09F9O01	rickettsia
11	236.5	28.4	131	2	052637	rickettsia
12	235.5	28.3	131	2	091522	rickettsia
13	216.5	26.0	105	2	031208	rickettsia
14	139	16.7	77	2	09AGC7	rickettsia
15	137	16.5	199	16	0985G4	rhizobium
16	127.5	15.3	182	16	09HX13	pseudomonas

17	120.5	14.5	136	16	092R89	092r89 rhizobium m
18	111.5	13.4	154	16	0914S1	0914s1 pseudomonas
19	109	13.1	222	16	09A3X8	09a3x8 caulobacter
20	109	13.1	237	16	09PGX0	09pgx0 xyliella fas
21	108.5	13.0	155	2	09F6B1	09f6b1 edwardsiella
22	108.5	13.0	221	16	092ST9	092st9 rhizobium m
23	106.5	12.8	155	2	09RA95	09ra95 serralia sp
24	105	12.6	155	2	09RB08	09rb08 pectobacter
25	104	12.5	139	2	052854	052854 rhizobium l
26	102.5	12.3	223	2	09XC44	09xc44 porphyromon
27	100.5	12.1	153	2	069776	069776 rhizobium e
28	100.5	12.1	257	16	09A8M8	09a8m8 caulobacter
29	99.5	12.0	83	16	092LP2	092lp2 rhizobium m
30	99.5	12.0	304	16	092762	092762 pseudomonas
31	99	11.9	332	16	09A9X2	09a9x2 pseudomonas
32	98	11.8	154	16	09CN83	09cn83 caulobacter
33	97	11.7	220	2	09XAX8	09xax8 pseudomonas
34	97	11.7	223	16	09KSR1	09ksr1 vibrio chol
35	95.5	11.5	79	16	092NT4	092nt4 escherichia
36	95	11.4	172	16	P76572	P76572 escherichia
37	94	11.3	608	10	09SUX1	09sux1 arabidopsis
38	93.5	11.2	838	2	09AL49	09al49 shigella fl
39	93	11.2	105	16	098P93	098p93 rhizobium l
40	91	10.9	105	16	0983Y0	0983y0 rhizobium l
41	90	10.8	137	16	09HU07	09hu07 pseudomonas
42	89.5	10.8	242	16	09PGD9	09pgd9 xyliella fas
43	88	10.6	544	5	046171	046171 nephila cla
44	87.5	10.5	147	5	09VS43	09vs43 drosophila
45	86.5	10.4	99	16	09HWP0	09hwp0 pseudomonas

ALIGNMENTS

RESULT 1	ID	Q9F9K8	PRELIMINARY:	PRT:	162 AA.
AC	Q9F9K8	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)				
DE	17 KDA ANTIGEN.				
GN	OSPA.				
OS	Piscirickettsia salmonis.				
OC	Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;				
OX	NCBI_TaxID=1238;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-LF-89;				
RA	Kuz'yk M.A., Burian J., Thornton J.C., Kay W.W.;				
RT	"Identification of a genus-common Rickettsial surface antigen in the				
RT	salmonid pathogen Piscirickettsia salmonis.";				
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AF184152; AAC17000.1; -				
SO	SEQUENCE 162 AA; 17661 MW; DDE99E6FD9A4527E CRC64;				

Query Match 98.0%; Score 815; DB 2; Length 162;  
Best Local Similarity 98.8%; Pred. No. 1.5e-65;  
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	2	RGCLQGSSLLIITSVFLVGCANFSPROEVGAATGAVVGAGVQGLFGKSGSRVNAIGAVL 61	
DB	3	RGCLQGSSLLIITSVFLVGCANFSPROEVGAATGAVVGAGVQGLFGKSGSRVNAIGAVL 62	
QY	62	GGILGSKIGSMPOODKIKLNQSLSEKAKAOVTRMRPDTGNSYSPVPTTYGRNKOER 121	
DB	63	GGILGSKIGSMPOODKIKLNQSLSEKAKAOVTRMRPDTGNSYSPVPTTYGRNKOER 122	
QY	122	RGQYCFREFQOKAMITAGOKOEIYGTACPPDGRMOWISTEK 161	
DB	123	RGQYCFREFQOKAMITAGOKOEIYGTACPPDGRMOWISTEK 162	

## RESULT 2

054381 ID 054381 PRELIMINARY; PRT: 148 AA.  
 AC 054381: 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE 17 KDA COMMON-ANTIGEN (FRAGMENT).  
 OS Rickettsia sp.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RX NCBI\_TaxID=789;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Davis M.J., Ying Z., Brunner B.R., Pantloja A., Ferwerda F.H.;  
 RT "Rickettsial relative associated with papaya bunchy top disease."  
 RL Curr. Microbiol. 36:80-84(1998).  
 DR EMBL; U76907; AAC02809.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 148 AA; 15050 MW; A7AFEEFDE0AE84C CRC64;

## Query Match

Best Local Similarity 36.5%; Score 303.5; DB 2; Length 148;  
 Matches 55; Conservative 30; Mismatches 43; Indels 9; Gaps 2;

QY 25 SROEVAATGAVVGGVAGGGLFEKSGSRVSMAGAVLGLGSLKIGSGMDQDK---IK 80  
 DB 17 NKGGSGTLTGTLGGVGLGSGGTGRLAAGAGALLGAILGAGMDQDKLAELT 76  
 QY 81 LKNSLEKAVAGVTRMRNPDGNSYSVEPVRYQRNKGROOYCFEFOQKAMINGO 140  
 DB 77 SORALEAPSGSSVGMNDNGNYGTVPKATY-----KNTGQYCREYQTQVWVGKQO 131  
 QY 141 ELYGTACPOPGRMQVI 157  
 DB 132 KAYGTACRQPDGCMQV 148

## RESULT 3

09F9F2 ID 09F9F2 PRELIMINARY; PRT: 159 AA.  
 AC 09F9F2: 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE 17 KDA GENUS-COMMON ANTIGEN.  
 OS Rickettsia felis (Rickettsia azadi).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RX NCBI\_TaxID=42862;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21217364; PubMed=11321078;  
 RA Bouyer D.H., Stenos J., Croquet-Valdes P., Moron C.G., Popov V.L.,  
 RA Zavala-Velazquez J.E., Foil L.D., Stothard D.R., Azad A.F.;  
 RT "Rickettsia felis: molecular characterization of a new member of the  
 RT spotted fever group."  
 RL Int. J. Syst. Evol. Microbiol. 51:339-347(2001).  
 DR EMBL; AF195118; AAG28452.1; -.  
 SQ SEQUENCE 159 AA; 16497 MW; 34C5B020AF470A1F CRC64;

## Query Match

Best Local Similarity 34.0%; Score 283; DB 2; Length 159;  
 Matches 61; Conservative 30; Mismatches 54; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AQNFSEVGAATGAVVGGVAGGGLFEKSGSRVSMAG-CAVL 61  
 DB 5 SKIITIALASMLQACNGCGMKNKQGTGTLGGAGAGALLGSGQFGKQQL-VGVGVALL 63

QY 62 GGLIGSKIGSGMDQDK---IKLNSLEKAVAGVTRMRNPDGNSYSVEPVRYQRN 117  
 DB 64 GAVLGGLIGAGMDQDKRIALTSORALEATPSGVSERNPDNGNHVYTPKTY----- 119  
 QY 118 KQERROOYCFEFOQKAMINGOKEIYGTACPOPDGRMOVIS 158  
 DB 120 -RNSTGQYCREYQTQVWVGKQOKAVGNCRQPDGLMOVN 159

## RESULT 4

052252 ID 052252 PRELIMINARY; PRT: 137 AA.  
 AC 052252: 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE 17 KDA ANTIGEN (FRAGMENT).  
 OS Rickettsia coolleyi.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RX NCBI\_TaxID=69410;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Billings A.N., Tellow G.J., Walker D.H.;  
 RT "Molecular characterization of a novel spotted fever group rickettsia  
 RT species from Ixodes scapularis in Texas."  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF031534; AAB95267.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 137 AA; 14215 MW; A27597A9AFD85FC3 CRC64;

## Query Match

Best Local Similarity 31.1%; Score 258.5; DB 2; Length 137;  
 Matches 53; Conservative 25; Mismatches 46; Indels 11; Gaps 4;

QY 22 QNFSEVGAATGAVVGGVAGGGLFEKSGSRVSMAG-CAVLGGLIGSKIGSGMDQDK-- 78  
 DB 7 RGMNKGSTLTGAGAGALLGSGQFGKQQL-VGVGVALLGAVLGGLIGAGMDQDKRL 65  
 QY 79 --IKLNSLEKAVAGVTRMRNPDGNSYSVEPVRYQRNKGROOYCFEFOQKAMIA 136  
 DB 66 AELTSORALEAPSGSSVEMRNPDNGNYGTVPKTY-----RNSTGQYCREYQTQVWVG 120  
 QY 137 GQKQIYGTACPOPD 151  
 DB 121 GKQOKAYGNACRQPD 135

## RESULT 5

09K2N6 ID 09K2N6 PRELIMINARY; PRT: 144 AA.  
 AC 09K2N6: 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE 17KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).  
 OS male-killing Rickettsia from Adalia bipunctata.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RX NCBI\_TaxID=38028;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schlenker H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,  
 RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;  
 RT "On the evolution of male-killing: Monophyletic origin and horizontal  
 RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-  
 RT generic ladybirds, Adalia bipunctata L. and A. decempunctata L.  
 RT (Coleoptera: Coccinellidae)."  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ269518; CAB96383.1; -.  
 DR EMBL; AJ269517; CAB96382.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 144 AA; 144

```

DE 17 KDA ANTIGEN (FRAGMENT).
OS male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiense; Rickettsia.
RN NCBI_TaxID=120393;
[1]
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=20575219; PubMed=11133455;
RA Schlenker D., Hurd G.D., Habig M., Sloggett J.J., Webberley M.K.,
RA Berland D., Hurd G.D., Maternus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot Ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT Coccinellidae).";
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL; AJ269516; CAB96381.1; -.
FT NON-TER 1
FT 1
SQ SEQUENCE 144 AA; 14801 MW; C825472F16A56AE7 CRC64;

Query Match
Best Local Similarity 30.2%; Score 251.5; DB 2; Length 144;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4;

QY 25 SROEVGATGAVVGAGOLFEKSGSRVSMAG-GAVLGLGKISIKGSDMOODK----I 79
DB 17 NKQGTGTLTGAGAGALLGSQFGKRGQL-VGVGVALLGAVLIGGIGAGMDODRRLAEL 75
QY 80 KLNQSLKKKAAQVVRMRNPDGTGNSYSEVPVTRYRNKQNEEROQYCRFQOKAMIAOK 139
DB 76 TISORLEAPSSSNVEMRNPDGNGHYTPNRTY-----RNSGTQCYREYTOTVVIAGKQ 130
QY 140 OEYGTACPOP 151
DB 131 QKSYGNACQPD 142

RESULT 8
O53154 PRELIMINARY; PRT; 154 AA.
ID AC O53154;
DT 01-NOV-1996 (TREMBLrel. 01. Created)
DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08. Last annotation update)
DE (CLONE PRB F1SF 1), 5' END CDS (FRAGMENT).
OS Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiense; Rickettsia.
OX NCBI_TaxID=789;
[1]
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=93084757; PubMed=1452660;
RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
RT "Characterization and comparison of Australian human spotted fever
RT group rickettsiae.";
RT J. Clin. Microbiol. 30:2896-2902(1992).
DR EMBL; M99391; AAA73386.1; -.
FT NON-TER 154
FT 154
SQ SEQUENCE 154 AA; 15849 MW; F5C35855EDBA439D2 CRC64;

Query Match
Best Local Similarity 29.3%; Score 244; DB 2; Length 154;
Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

QY 8 SLLIITV---FLVGC--AQNFSRQEVGAATGAVVGAGVAGOLFEGKSGSRVSMAG-GAVL 61
DB 5 SKIMITALATSMLOACNGPGGNKQGTGTLTGAGAGALLGSQFGKRGQL-VGVGVALL 63
QY 62 GLGISKSIKIGSDMOODK----IKLNQSLKKKAAQVVRMRNPDGTGNSYSEVPVTRYORN 117
DB 64 GAVLGGQIGAGMDODRRLAELTISORALETPADSGSNVEMRNPDGNSYGYVTPNKRTRNST 123
QY 118 KOERROQYCRFQOKAMIAQOKOEIYGTACPOPGR 153

```

Db 124 G0D-----CRVYTOTVIVIGSKQKQKAYGNACRQPDGQ 154

## RESULT 9

Q9F909 ID PRELIMINARY; PRT; 151 AA.  
AC 09F909: 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
OS OUTER MEMBRANE PROTEIN (FRAGMENT).  
OC Rickettsia heidelberg.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
RN NCBI\_Taxid=35789;  
RP SEQUENCE FROM N.A.  
RA Nilsson K., Pahlson C.;  
RT "Novel peptide diagnostic reagent and kit for detection of  
Rickettsiosis."  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF181036; AAG09427.1; -.  
FT NON\_TER 151  
SQ SEQUENCE 151 AA: 15621 MW: 877407B9C71E4B39 CRC64;

## Query Match

Best Local Similarity 28.4%; Score 239; DB 2; Length 151;  
Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

OY 8 SLLIISV---FLVGC--AQNRSREVGAAATGAVGAGOLFEGSGSRVSMALIG-GAVL 61  
Db 5 SKIMITALAAMQACNGPBGKMKGTGTLGGAGGALLSQFGKGGQL-VGVGVAL 63  
OY 62 GGLIGSKIGSMDOODK---IKLNQSLKVKAGOVTRMNPPTGNSYSEVPYRYRYN 117  
Db 64 GAVLGGQVAGAGDEDRRLAELTQSRLAEAPSSNVEMNPDMNGYVTPNKTY----- 119  
OY 118 KOERROQYCRFEOOKAMIAIGOKEIYGTACPOP 150  
Db 120 -RNSTGYCREYTOTVIVIGSKQKQKAYGNACRQPDG 151

## RESULT 10

O9F001 ID PRELIMINARY; PRT; 131 AA.  
AC 09F001: 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
OS 17 KDA PROTEIN (FRAGMENT).  
OC Rickettsia sp. California 2.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
RN NCBI\_Taxid=147259;  
RP SEQUENCE FROM N.A.  
RA Raoult D.;  
RT "A new SFG rickettsia isolated from fleas."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [12]  
RP SEQUENCE FROM N.A.  
RA Roux V., Raoult D.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF106593; AAC48554.1; -.  
FT NON\_TER 131  
SQ SEQUENCE 131 AA: 13374 MW: 2308819B29FF860 CRC64;

## Query Match

Best Local Similarity 28.4%; Score 236.5; DB 2; Length 131;  
Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

OY 25 SROEVGATGAVGVAGOLFEGSGSRVSMALIG-GAVLGGILGSKIGSMDOODK-----I 79  
Db 10 NKQGTGTLGAGAGALLGSGFGKGGQL-VGVGVALGAVAGGIGAGMEDODRLAEL 68  
OY 80 KLNQSLKVKAGOVTRMNPPTGNSYSEVPYRYRYNKOERROQYCRFEOOKAMIAIGOK 139  
Db 69 TSOALAEAPSSNVEMNPDMNGYVTPNKTY-----RNSTGYCREYTOTVIVIGGQ 123  
OY 140 QEITYGTAC 147  
Db 124 QKAYGNAC 131

## RESULT 11

O52637 ID PRELIMINARY; PRT; 131 AA.  
AC 052637: 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE 01-NOV-1996 (TREMBLrel. 08, Last annotation update)  
OS Rickettsia sp.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
RN NCBI\_Taxid=789;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=94117373; PubMed=8288533;  
RA Warren J.H., Hurst G.D., Zhang W., Breeuwer J.A., Scouthamer R.,  
RA Majerus W.E.;  
RT "Rickettsial relative associated with male killing in the ladybird  
beetle (Adalia bipunctata)."  
RL J. Bacteriol. 176:388-394(1994).  
DR EMBL; U04162; AAA19235.1; -.  
FT NON\_TER 131  
SQ SEQUENCE 131 AA: 13344 MW: A1DCFF71050DF52DF CRC64;

## Query Match

Best Local Similarity 28.4%; Score 236.5; DB 2; Length 131;  
Matches 49; Conservative 24; Mismatches 44; Indels 11; Gaps 4;

OY 25 SROEVGATGAVGVAGOLFEGSGSRVSMALIG-GAVLGGILGSKIGSMDOODK-----I 79  
Db 10 NKQGTGTLGAGAGALLGSGFGKGGQL-VGVGVALGAVAGGIGAGMEDODRLAEL 68  
OY 80 KLNQSLKVKAGOVTRMNPPTGNSYSEVPYRYRYNKOERROQYCRFEOOKAMIAIGOK 139  
Db 69 TSOALAEAPSSNVEMNPDMNGYVTPNKTY-----RNSTGYCREYTOTVIVIGGQ 123  
OY 140 QEITYGTAC 147  
Db 124 QKAYGNAC 131

## RESULT 12

O9L522 ID PRELIMINARY; PRT; 131 AA.  
AC 09L522: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
OS 17 KDA SURFACE ANTIGEN (FRAGMENT).  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
RN NCBI\_Taxid=47589;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=21091941; PubMed=11157215;  
RA Simser J.A., Palmer A.T., Munderloh U.G., Kurtti T.J.;





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein . protein search, using sw model

Run on: October 27, 2002, 11:00:36 : Search time 6.11744 Seconds

(without alignments)  
1019.028 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832  
Sequence: 1 MRGCLQGSSLIITVFLVGC.....IYCTACPDPGRMQVISTEK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SWISSProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	288	34.6	159	17KD_RICPR	P16624 rickettsia
2	288	34.4	159	17KD_RICJA	Q52764 rickettsia
3	285	34.3	159	17KD_RICCN	P05372 rickettsia
4	276.5	33.2	159	17KD_RICTY	P22882 rickettsia
5	268	32.2	154	17KD_RICAU	P50928 rickettsia
6	259	31.1	154	17KD_RICPA	P50930 rickettsia
7	259	31.1	154	17KD_RICRH	P50931 rickettsia
8	256	30.8	154	17KD_RICMO	P50929 rickettsia
9	255	30.6	154	17KD_RICAM	P50927 rickettsia
10	156	18.8	80	17KD_RICCA	P29697 rickettsia
11	112.5	13.5	155	PCP_YEREN	P31464 yersinia en
12	105.5	12.7	155	SLVB_SALTY	Q33349 salmonella
13	104.5	12.6	155	SLVB_ECOLI	P55741 escherichia
14	101	12.1	155	PCP_HAEIN	P10335 haemophilus
15	100	12.0	179	YCFJ_ECOLI	P37796 escherichia
16	90	10.8	526	KICJ_BOVIN	P06394 bos taurus
17	90	10.8	1332	KKDO_BACSU	P45334 bacillus su
18	88	10.6	1585	YOBQ_BACSU	P45931 bacillus su
19	86.5	10.4	72	OSMB_SALTY	P37723 salmonella
20	85.5	10.3	72	OSMB_ECOLI	P45931 escherichia
21	84.5	10.2	431	KRE2_CANAL	Q00310 candida alb
22	82.5	9.9	243	CYSH_SALTY	P17853 salmonella
23	82.5	9.9	541	NU57_YEAST	P48837 saccharomyc
24	82	9.9	132	V615_AOUAE	O66867 aquilex aeo
25	81.5	9.8	301	STXG_RAT	Q92158 rattus norv
26	81.5	9.8	526	VP5_BTIV1	P33476 bluetongue
27	80	9.6	806	ITB7_MOUSE	P26011 mus musculu
28	79.5	9.6	113	YKR3_MOUSE	P4309 caenorhabdi
29	79.5	9.6	263	CANS_BOVIN	P13135 bos taurus
30	79.5	9.5	593	KICJ_HUMAN	P13045 homo sapien
31	78.5	9.4	219	YIAD_ECOLI	P37665 escherichia
32	78.5	9.4	747	SPDI_NEPCCL	P19837 nephila cla
33	78	9.4	514	ATPA_THIFE	P41167 thiodacillu

## ALIGNMENTS

34	77.5	9.3	359	1	ATPA_BOVIN	P19482 bos taurus
35	77.5	9.3	467	1	HEMI_MYCLE	P46774 mycobacteri
36	77.5	9.3	543	1	ATPA_RAT	P15999 rattus norv
37	77.5	9.3	553	1	ATPA_HUMAN	P25705 homo sapien
38	77	9.3	266	1	CANS_RABIT	P06813 oryctolagus
39	76.5	9.2	553	1	ATPO_BOVIN	P19483 bos taurus
40	76.5	9.2	553	1	ATPA_MOUSE	O03265 mus musculu
41	76.5	9.2	569	1	KICJ_MOUSE	P02535 mus musculu
42	76.5	9.2	727	1	IF2M_HUMAN	P46189 homo sapien
43	76.5	9.2	747	1	EL5_BOVIN	P04985 bos taurus
44	76.5	9.2	2516	1	ECAD_DROME	O24270 drosophila
45	76	9.1	526	1	VP5_BTIV10	P07389 bluetongue

RESULT 1  
17KD\_RICPR STANDARD; PRT; 159 AA.

AC P16624;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 17 kda surface antigen precursor.

GN OMP OR RP833.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiense; Rickettsia.  
OX NCBI\_Taxid=782;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-MADRID E.  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Slichter-Ponten T., Alsmark U.C.M., Pedowski R.M., Naeslund A.K.,  
RT Eriksson A.S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria."  
RT Nature 396:133-140(1998).  
RL Nature 396:133-140(1998).  
RT Comparative sequence analysis of a genus-common rickettsial antigen gene."  
RT J. Bacteriol. 171:5199-5201(1989).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-MADRID E.  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Slichter-Ponten T., Alsmark U.C.M., Pedowski R.M., Naeslund A.K.,  
RT Eriksson A.S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria."  
RT Nature 396:133-140(1998).  
RL Nature 396:133-140(1998).  
RT Comparative sequence analysis of a genus-common rickettsial antigen gene."  
RT J. Bacteriol. 171:5199-5201(1989).  
RN [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M28482; AAA26378.1; ALT\_SRO.  
DR EMBL: AJ235273; CAA15258.1; -.  
DR PIR: D33971; D33971.  
DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.  
FT SIGNAL 1 19  
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.  
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
SQ SEQUENCE 159 AA; 16672 MW; A33D404B5EBB071 CRC64;

Query Match 34.6%; Score 288; DB 1; Length 159;  
Best Local Similarity 37.9%; Pred. No. 3; Le-18;  
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

```

QY 8 SLLIITSV---FLVGC--AQNFSROEVGATGAVVGGVAGOLFEKSGSRVSMAG-GAVL 61
Db 5 SKIMITALASMLQACNGSQSMNKGCTITLLGAGGLLSGFCQGGQGV-VGVGVALI 63
QY 62 GGLISKIKSGSDMOODK-----IKNLSLEKVAQVTRMNPDTGNSVSEVPRVYQRVN 117
Db 64 GAVLGGQIGASMDQDRRLLELTSORALESPSGNSNEMRNPNGHGVYTPKRTY----- 119
QY 118 KOERROQYCREFOQKAMIAQOKETIYGTACPOPDGRWQYIS 158
Db 120 -RNSAGQYCREYQTQTVILGGKQKQKGTGNACRQPDGQMOVYN 159

```

## RESULT 2

```

17KD_RICJA STANDARD: PRT: 159 AA.
AC 052764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 17 kDa surface antigen precursor.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OX Rickettsiaceae; Rickettsieae; Rickettsia.
NCBI_TaxId=35790;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RX MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
RT specimens by PCR.";
RL J. Clin. Microbiol. 33:487-489(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D16515; BAA03965.1; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 N-ACYL DIGLYCERIDE (PROBABLE).
FT LIPID 20 20
SQ SEQUENCE 159 AA; 16554 MW; CDDCE7CEBDCD6B41 CRC64;

```

```

Query Match 34.4%; Score 286; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 4; Ee-18;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;

```

```

QY 8 SLLIITSV---FLVGC--AQNFSROEVGATGAVVGGVAGOLFEKSGSRVSMAG-GAVL 61
Db 5 SKIMITALASMLQACNGSQSMNKGCTITLLGAGGLLSGFCQGGQGV-VGVGVALI 63
QY 62 GGLISKIKSGSDMOODK-----IKNLSLEKVAQVTRMNPDTGNSVSEVPRVYQRVN 117
Db 64 GAVLGGQIGASMDQDRRLLELTSORALESPSGNSNEMRNPNGHGVYTPKRTY----- 119
QY 118 KOERROQYCREFOQKAMIAQOKETIYGTACPOPDGRWQYIS 158
Db 120 -RNSAGQYCREYQTQTVILGGKQKQKGTGNACRQPDGQMOVYN 159

```

## RESULT 3

```

17KD_RICCN STANDARD: PRT: 159 AA.
ID 17KD_RICCN

```

```

AC P05372;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RCL1287.
OS Rickettsia conorii, and
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OX Rickettsiaceae; Rickettsieae; Rickettsia.
NCBI_TaxId=781, 783;
RN (1)
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii, and R.rickettsii;
RX MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN (2)
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii; STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN (3)
RP SEQUENCE FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=87222152; PubMed=3108232;
RA Anderson B.E., Regnier R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA Fu Z.Y., Bellini W.J.;
RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
RT rickettsii.";
RL J. Bacteriol. 169:2385-2390(1987).
RN (4)
RP SEQUENCE OF 1-30 FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=89008059; PubMed=3139629;
RA Anderson B.E., Baumstark B.R., Bellini W.J.;
RT "Expression of the gene encoding the 17-kilodalton antigen from
RT Rickettsia rickettsii: transcription and posttranslational
RT modification.";
RL J. Bacteriol. 170:4493-4500(1988).
RN (5)
RP SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
RN anchor (Probable).
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@sib-sib.ch).
RN -----
DR EMBL; M28479; AAA26379.1; -
DR EMBL; M28480; AAA26376.1; -
DR EMBL; AE008675; AAL03825.1; -
DR EMBL; M16486; AAA26381.1; -
DR EMBL; J03371; -; NOT_ANNOTATED_CDS.
DR PIR; A25972; A25972.
DR PIR; A31836; A31836.
DR PIR; A33971; A33971.
DR PIR; B33971; B33971.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159
FT LIPID 20 20
FT CONFLICT 146 146 N-ACYL DIGLYCERIDE (PROBABLE);
FT CONFLICT 153 153 G->E (IN REF. 3);

```

```

Db      141 QTYGNACRQDPDGMQVYN 159

RESULT 5
17KD_RICAU
ID      17KD_RICAU      STANDARD:      PRT:      154 AA.
AC      P50928;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
OMP.
OS      Rickettsia australis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxID=787;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Baird R.W., Ross B., Dwyer B.;
RL      Submitted (OCT-1991) to the EMBL/Genbank/DBJ databases.
CC      -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC      anchor (Probable).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.sib.ch).
-----
DR      EMBL; M74042; AAA26394.1; '-'
DC      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Outer membrane; Lipoprotein; Antigen; Signal.
FT      SIGNAL 1 19 BY SIMILARITY.
FT      CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT      LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT      NON_TER 154 154
SQ      SEQUENCE 154 AA; 15967 MW; E3AA833346FAC320 CRC64;

Query Match 32.2%; Score 268; DB 1; Length 154;
Best Local Similarity 37.8%; Pred. No. 1.6e-16;
Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;

QY      8 SLLIIIV---FLVGCAG--NFSRQEVGATGAVVGVAGQIFGKSGSRVSMATG-GAVL 61
      I::I::: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5 SKIMIALAASMLQACNSPGGMKNKGCTGTLLGAGGALLGSGFGKGGQL-VGVGVALL 63

QY      62 GGLIGSKTGQSDQODK---IKLNQSLLEKKYKAGQVYTRMRNDPTDGNSTSVPEVFTYQRYN 117
      I::I::I | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      64 GAVLIGAGAGAGDEDDRLRLAELTSGRALETAPASGVNEMRDNDCNGYVTPNKKTYRMSN 123

QY      118 KQERRQGYCREFGQKAMTAGOKQELTYGNACPODPR 153
      I::I::: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      124 ----GQYCREYQTQTVIGRKQKAYGNACRQPDQ 154

RESULT 6
17KD_RICPA
ID      17KD_RICPA      STANDARD:      PRT:      154 AA.
AC      P50930;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      17 kDa surface antigen precursor (Fragment).
OMP.
OS      Rickettsia parkeri.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiae; Rickettsia.
OX      NCBI_TaxID=35792;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strain=MACULATUM.
RC

```

```

RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11020; AAB07706.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154
FT LIPID 20 20
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;

Query Match
Best Local Similarity 31.1%; Score 259; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AONFSROEVGATGAVGVAGOLFSGSGRVSMATG-GAVL 61
D 5 SKIMVIALASMLAQCNGPGMKNKGCTGLLGAGGALLGSGFGKGGQL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNSLEKFKAGQVTRMRNPDTGNSYSEVPRTYQRN 117
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

Db 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

RESULT 7
17KD-RICRH
ID 17KD-RICRH STANDARD; PRT; 154 AA.
AC P50931;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia rhipicephali.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33992;
RN [1]
RP SEQUENCE FROM N.A.
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.
CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11020; AAB07706.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154
FT LIPID 20 20
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9DBD5EEC CRC64;

Query Match
Best Local Similarity 30.8%; Score 256; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AONFSROEVGATGAVGVAGOLFSGSGRVSMATG-GAVL 61
D 5 SKIMVIALASMLAQCNGPGMKNKGCTGLLGAGGALLGSGFGKGGQL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNSLEKFKAGQVTRMRNPDTGNSYSEVPRTYQRN 117
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

Db 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

```

```

FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15895 MW; 0CF85AD5D96DFEFB CRC64;

Query Match
Best Local Similarity 31.1%; Score 259; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AONFSROEVGATGAVGVAGOLFSGSGRVSMATG-GAVL 61
D 5 SKIMVIALASMLAQCNGPGMKNKGCTGLLGAGGALLGSGFGKGGQL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNSLEKFKAGQVTRMRNPDTGNSYSEVPRTYQRN 117
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

Db 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

RESULT 8
17KD-RICMO
ID 17KD-RICMO STANDARD; PRT; 154 AA.
AC P50929;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=33991;
RN [1]
RP SEQUENCE FROM N.A.
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
CC - SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11017; AAB07705.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154
FT LIPID 20 20
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15881 MW; A09C53B8769E31DA CRC64;

Query Match
Best Local Similarity 30.8%; Score 256; DB 1; Length 154;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 8 SLLIISV---FLVGC--AONFSROEVGATGAVGVAGOLFSGSGRVSMATG-GAVL 61
D 5 SKIMVIALASMLAQCNGPGMKNKGCTGLLGAGGALLGSGFGKGGQL-VGVGVALL 63
QY 62 GGLIGSKIGSMDQDK-----IKLNSLEKFKAGQVTRMRNPDTGNSYSEVPRTYQRN 117
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 64 GAVLGQIGAGMDEQDRRLAELTISQRALETAPSGSNVEMRNPDNGNYITPNKTY----- 119
QY 118 KQERRQOYCREFOQKAMTAGQKQEIYGTACPOPDR 153
D 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

Db 120 -RNSTGYCREYQTVVIGKQKQKAYGNACIQPDQ 154

```

```

RESULT 9
17KD_RICAM STANDARD; PRT; 154 AA.
ID 17KD_RICAM
AC P50927;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia amblyomni1.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=33989;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO 85-1084;
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U11013; AAB07704.1; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 >154 BY SIMILARITY.
FT LIPID 20 20 17 KDA SURFACE ANTIGEN.
FT NON_TER 154 154 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 154 AA; 15879 MW; E4FBE4C29D943581 CRC64;

Query Match 30.6%; Score 255; DB 1; Length 154;
Best Local Similarity 36.5%; Pred. No. 2; Le-15;
Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

QY 8 SCLIIIV---FLVGC--AQNFSKQEVGAATGAVVGVAGOLGFGSGGRVSMATG-CAVL 61
DB 5 SKIMIALASTLOACNGPGMKNQGTGLTGAGALLGSGFGKGGQL-VGVGVGALL 63
QY 62 GGLGSGKIGOSMDQDK---IKLNSLEKVKACQVTRWRNPDGNSYSVEPVRTYQRYN 117
DB 64 GAVVGGVAGAMDQDRIRIELTSQKALETPAPNCSNVEWRNPNGNNGYTPNKTY---- 119
QY 118 KQERRQYCRERFQOKAMIAQKOEIVGTACPPDGR 153
DB 120 -RNSTGYCRHYTQTVVIGGKQKAVGNACRQPDGQ 154

RESULT 10
17KD_RICCA STANDARD; PRT; 80 AA.
ID 17KD_RICCA
AC P29697;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen (Fragment).
GN OMP.
OS Rickettsia canada.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxID=788;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92108069; Pubmed=1729713;

```

```

RA Azad A.F., Sacchi J.B. Jr., Nelson W.M., Dasch G.A.,
RA Schmidtmann E.T., Carl M.;
RT "Genetic characterization and transovarial transmission of a
RT typhus-like rickettsia found in cat fleas.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:43-46(1992).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M82879; -. NOT_ANNOTATED_CDS.
DR InterPro: IPR000437; PROK_LIPOPROT.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Outer membrane; Lipoprotein; Antigen.
FT NON_TER 1 1
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 8372 MW; AD289A48EAB19E0E CRC64;

Query Match 18.8%; Score 156; DB 1; Length 80;
Best Local Similarity 40.3%; Pred. No. 3; 9e-07;
Matches 31; Conservative 15; Mismatches 27; Indels 4; Gaps 1;

QY 42 GOLFSGSGRYSMAGVAVGLGSLGKIGOSMDQDK---IKLNSLEKVKACQVTRWR 97
DB 1 GSGFGKGGKGLIGVAGALLGAILGNQIGAGMDQDRRLAEVTSQRLAEVTSSTSEWR 60
QY 98 NPDGNSYSVEPVRTYQ 114
DB 61 NPDGNSYGYTPSKTYK 77

RESULT 11
PCP_YEREN STANDARD; PRT; 155 AA.
ID PCP_YEREN
AC P31484;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane lipoprotein pcv precursor.
GN PCP OR PCPY.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / SEROTYPE O:8;
RX MEDLINE=92121089; Pubmed=1732192;
RA Baumann A.J., Hantke K.;
RT "A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine
RT uptake in Escherichia coli.";
RL J. Bacteriol. 174:1029-1035(1992).
CC -1- FUNCTION: FACILITATES FERRIOXAMINE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO E. COLI AND S. TYPHIMURIO SLXB AND TO
CC H. INFLUENZAE PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60448; CAA42977.1; -.

```

```

DR PIR: S23787; S23787.
DR HSSP: P00778; 1P04.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN PCP.
FT LIPID 18 18 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 155 AA; 15362 MW; 8AD6BE2132E849FA CRC64;

Query Match 13.5%; Score 112.5; DB 1; Length 155;
Best Local Similarity 24.7%; Pred. No. 0.0044;
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;

QY 9 SLIISVFLVCCAGN-----FSRQE-----VGAAAT 33
DB 7 AVAIAAATLTGCAANNNTLSGDFVSAQAKOVTVTGTLISVPTTIGGGDDNNVGAIG 66
QY 34 GAVGVAGOLFGRGSRVMAIGAVLGLGSKIGSDQDQIKL----- 81
DB 67 GAVLGGFLGNTVGGGRSLATAGAVAGMGAGVGAMNRTDGVLEVRKDDGTTTLV 126
QY 82 --NQSLEKVKAGVYTRWRPDTGNSYSVEP 109
DB 127 VQKQGPTRFVSQ--RVMLASSGSTVTVP 154

RESULT 12
SLYB_SALTY
ID SLYB_SALTY STANDARD: PRT; 155 AA.
AC Q53549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
GN Outer membrane lipoprotein slyB precursor.
OS SLYB OR STM1445 OR STY1677.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OC Salmonella.
ON NCBI_TaxID=602, 601;
RX NCBI_TaxID=602, 601;
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=96133688; PubMed=8544813;
RA Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.;
RT "slyB, a regulatory protein from Salmonella typhimurium, induces a
RT haemolytic and pore-forming protein in Escherichia coli.";
RL Mol. Gen. Genet. 249:474-486(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies K.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,
RA Krogh A., Larsen T.S., Leather A., Hien T.T., Holroyd S., Jorgels K.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;

```

---

```

RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: TO E.COLI SLYB, H.INFLUENZAE PCP AND Y.ENTEROCOLITICA
CC PCP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isdb-sib.ch/announce/
CC or send an email to license@isdb-sib.ch).
CC -----
DR EMBL: S80790; AMB35871.2; -
DR EMBL: AE008762; AAL20367.1; -
DR EMBL: AL627271; CAD01922.1; -
DR STYGENE: SG10573; slyB.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Complete proteome.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 155 OUTER MEMBRANE LIPOPROTEIN SLYB.
FT LIPID 18 18 N-ACYL DIGLYCERIDE.
SQ SEQUENCE 155 AA; 15548 MW; 82FDDCDBABD5A7 CRC64;

Query Match 12.7%; Score 105.5; DB 1; Length 155;
Best Local Similarity 24.3%; Pred. No. 0.018;
Matches 34; Conservative 20; Mismatches 53; Indels 33; Gaps 3;

QY 1 MRGLGSSS-----IISVFLVCCAGNFSRQEVGAATGAVGVA 41
DB 15 LACGVNNDLSGCDVYASPAKQVNTTGTIVNVPVIOGGDSNVVIGAGVAVLGFL 74
QY 42 GOLFGKSGSRVMAIGAVLGLGSKIGSDQDQIKLQNSLEKVKAGVTRWRPDT 101
DB 75 GNTIGGTRSLATAGAVAGMGAGVGAMNRTDGVLEVRKDDGTTTLV 126
QY 102 GNSYSVEPYTRYRNRKOE 121
DB 121 GNTIMVQKQGPTRFVSQ--RVMLASSGSTVTVP 140

RESULT 13
SLYB_ECOLI
ID SLYB_ECOLI STANDARD: PRT; 155 AA.
AC P55741; P76183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Outer membrane lipoprotein slyB precursor.
OS SLYB OR BL641 OR Z2655 OR ECS2350.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OC Escherichia.
ON NCBI_TaxID=562, 83334;
RX NCBI_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=96133688; PubMed=8544813;
RA Ludwig A., Tengel C., Bauer S., Hubert A., Benz R., Mollenkopf H.-J.,
RA Goebel W.;
RT "slyB, a regulatory protein from Salmonella typhimurium, induces a
RT haemolytic and pore-forming protein in Escherichia coli.";
RL Mol. Gen. Genet. 249:474-486(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

```





Query Match 12.1%; Score 101; DB 1; Length 155;  
 Best Local Similarity 39.6%; Pred. No. 0.043;  
 Matches 21; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

QY 29 VCAATGAVGVAGVAGQLEFGKSGRVSMATGAVLGLGSKIGSMQDQDKIKL 81  
 DB 62 VETLGGALGAGTGTGGGKGAIAVGAIGATAGSKTEKMSQVNGAEL 114

RESULT 15  
 ID YCFJ\_ECOLI STANDARD; PRT; 179 AA.  
 AC P3796: P75951;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 GN Hypothetical protein ycfJ.  
 OS YCFJ OR B1110.  
 OC Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [3]  
 RP SEQUENCE OF 1-63 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=81236546; PubMed=6265208;  
 RA Young J.G., Rogers B.L., Campbell H.D., Jaworski A., Shaw D.C.;  
 RT "Nucleotide sequence coding for the respiratory NADH dehydrogenase of  
 Escherichia coli. UUG initiation codon.";  
 RL Eur. J. Biochem. 116:165-170(1981).  
 RN [4]  
 RP IDENTIFICATION.  
 RX MEDLINE=95075659; PubMed=7984428;  
 RA Borodovsky M., Rudd K.E., Koonin E.V.;  
 RT "Intrinsic and extrinsic approaches for detecting genes in a  
 bacterial genome.";  
 RL Nucleic Acids Res. 22:4756-4767(1994).  
 CC -I- SIMILARITY: TO RICKETTSIA 17 KDA SURFACE ANTIGEN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 DR EMBL: A600021; AAC74194.1; -;  
 DR EMBL: D90746; BAA55925.1; -;  
 DR EMBL: V00306; -; NOT\_ANNOTATED\_CDS.

DR EcGene: E312444; ycfJ.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 5 POTENTIAL.  
 SQ SEQUENCE 179 AA; 18920 MW; BA5EB0DB56DA5609 CRC64;

Query Match 12.0%; Score 100; DB 1; Length 179;  
 Best Local Similarity 26.3%; Pred. No. 0.061;  
 Matches 36; Conservative 17; Mismatches 48; Indels 36; Gaps 4;

QY 30 GAATGAVGVAGVAGQLEFGKSGRVSMATGAVLGLGSKIGSMQDQDKIKL 89  
 DB 73 GSVLGAIVAGVAGVAGQLEFGKSGRVSMATGAVLGLGSKIGSMQDQDKIKL 120  
 QY 90 ACGVTRMRNPDTGNSVPEVRYQRYNKOERROQYCFEFQKAMTAGCKOEIY-----G 144  
 DB 121 -----TYTTTQQRCKTVYDKSKMGLGVYTK-----IGDQCKIRMDRDPG 162  
 QY 145 TACPDGGRWQVISTEK 161  
 DB 163 TQIRL-DSNGQLILNKK 178

Search completed: October 27, 2002, 11:10:51  
 Job time : 6.11744 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:05:56 ; Search time 10.2884 Seconds  
(without alignments)  
1503.668 Million cell updates/sec

Title: US-09-677-374-4

Perfect score: 832

Sequence: 1 MRGCGSSSLIITISVFLVGC.....IYGTACPDGDMQVISTEK 161

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	288	34.6	159	2 D33971	ricketsial common
2	285	34.3	159	2 B33971	ricketsial common
3	285	34.3	159	2 A33971	ricketsial common
4	285	34.3	159	2 G97860	17K surface antigen
5	276.5	33.2	159	2 C33971	ricketsial common
6	276	33.2	159	2 A25972	17K antigen precursor
7	127.5	15.3	182	2 D83169	conserved hypochet
8	120.5	14.5	131	2 A13418	17K surface antigen
9	112.5	13.4	155	2 S23787	outer membrane lip
10	111.5	13.4	154	2 B83514	conserved hypochet
11	109	13.1	179	2 AE0644	probable secreted
12	109	13.1	232	2 G87629	hypothetical prote
13	109	13.1	257	2 B82837	hypothetical prote
14	107.5	12.9	142	2 AD2696	lipa protein (limpo
15	107.5	12.9	155	2 AF0289	probable lipoprote
16	107	12.9	125	2 D97478	lipa protein (limpo
17	105.5	12.7	155	2 A10693	outer membrane lip
18	104.5	12.6	155	2 C64921	outer membrane lip
19	104.5	12.6	155	2 F90922	probable outer mem
20	104.5	12.6	155	2 B85771	probable outer mem
21	104	12.5	139	2 S58234	lipa protein - Rhl
22	103	12.4	155	2 AG0443	outer membrane lip
23	103	12.4	232	2 AD3350	outer membrane pro
24	101	12.1	155	2 T64130	hypothetical prote
25	100.5	12.1	257	2 F87413	hypothetical prote
26	100	12.0	179	2 C64855	ycyf protein - Esc
27	100	12.0	179	2 D85674	hypothetical prote
28	100	12.0	179	2 H90814	hypothetical prote
29	99.5	12.0	304	2 H83636	hypothetical prote

30	99	11.9	332	2 D87353	hypothetical prote
31	97	11.7	179	2 AC0198	probable exported
32	97	11.7	223	2 C82230	probable lipoprote
33	95	11.4	172	2 G91049	probable outer mem
34	95	11.4	172	2 H65026	hypothetical prote
35	95	11.4	172	2 D85894	probable outer mem
36	94	11.3	278	2 AB3091	hypothetical prote
37	94	11.3	278	2 H98195	hypothetical prote
38	94	11.3	608	2 T03442	glycine-rich prote
39	92.5	11.1	220	2 G97685	probable outer mem
40	92.5	11.1	220	2 A12910	porin (imported) -
41	92	11.1	100	2 G97672	hypothetical prote
42	92	11.1	100	2 AE2897	conserved hypochet
43	90.5	10.9	691	2 F91251	probable tape meas
44	90	10.8	137	2 B82998	hypothetical prote
45	90	10.8	526	1 KRBOVI	keratin, 54K type

ALIGNMENTS

```
RESULT 1
D33971
ricketsial common antigen precursor - Rickettsia prowazekii
N:Alternate names: 17K surface antigen; outer membrane protein (omp); RP833
C:Species: Rickettsia prowazekii
C>Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 03-Nov-2000
C:Accession: D33971; B71645
R:Anderson, B.E.; Tzianabos, T.
J. Bacteriol. 171, 5199-5201, 1989
A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.
A:Reference number: A33971; M0ID:89359171
A:Accession: D33971
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-159 <AND>
A:Cross-references: GB:M28482; NID:q152461
A:Note: the sequence in GenBank entry R18A117K, release 109.0, (PID:q152462) omits
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alismark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; M0ID:99039499
A:Accession: B71645
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-159 <AN2>
A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CA15258.1; PID:g386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: omp; RP833
C:Superfamily: rickettsial common antigen
C:Keywords: surface antigen

Query Match          34.6%; Score 288; DB 2; Length 159;
Best Local Similarity 37.9%; Pred. No. 1.7e-18;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
```

QY 8 SLLIISV---FLVGC--AQNFSRQEVGAATGVAGVAGQLFGKSGSRVMAIG-GAVL 61  
DB 5 SKIMITIALASMLQACAGSGGNKKGGTLLGGAGGALLSGQGQKGLQ-VGVGVALL 63

QY 62 GGLIGSKTIGSMQODK---IKLNSLEKVKAGQVTRMPNPTGNSYSEVPRTYRYN 117  
DB 64 GAVLGIGIGASMBEDPRRLLETLSORALLESAPGSNTEHNPNGNHYTPNKTY---- 119

QY 118 KOERROOYCREFOQKAMIAQOKOIEYGTACPOPDGRMOVTS 158  
DB 120 -RNSAGQYCREYOTYVIGSKQKRTYGNACRQPDGQMOVYN 159

RESULT 2  
B33971  
ricketsial common antigen precursor - Rickettsia conorii

C:Species: Rickettsia conorii  
 C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
 C:Accession: B33971  
 R:Anderson, B.E.; Tzianabos, T.  
 J. Bacteriol. 171, 5199-5201, 1989  
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
 A:Reference number: A33971; MUID:89359171  
 A:Accession: B33971  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <AND>  
 A:Cross-references: GB:M28479; NID:9152463; PIDN:AAA26379.1; PID:9152464  
 C:Superfamily: rickettsial common antigen

Query Match 34.3%; Score 285; DB 2; Length 159;  
 Best Local Similarity 37.9%; Pred. No. 3.1e-18;  
 Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;  
 QY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVL 61  
 D 5 SKIMIALATSMLOACNGPGMKNKGOTGTLTGAGAGALLGSGFGKKGQL-VGVGVAL 63  
 QY 62 GGLIGSKIGSMDOQDK---IKLNSLEKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYN 117  
 D 64 GAVLGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEMRNDNGNYGYVTENKTY---- 119  
 QY 118 KOERROQYCREFOOKAMIAQOKOETIGTACPOPDGRMOVIS 158  
 D 120 -RNSTGQYCRETYOTYVIGSKQOKAIGNACROPDQOMQVYN 159

RESULT 3  
 A33971  
 Rickettsial common antigen precursor - Rickettsia rickettsii  
 C:Species: Rickettsia rickettsii  
 C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
 C:Accession: A33971  
 R:Anderson, B.E.; Tzianabos, T.  
 J. Bacteriol. 171, 5199-5201, 1989  
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
 A:Reference number: A33971; MUID:89359171  
 A:Accession: A33971  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <AND>  
 A:Cross-references: GB:M28480; NID:9152457; PIDN:AAA26376.1; PID:9152458  
 C:Superfamily: rickettsial common antigen

Query Match 34.3%; Score 285; DB 2; Length 159;  
 Best Local Similarity 37.9%; Pred. No. 3.1e-18;  
 Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;  
 QY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVL 61  
 D 5 SKIMIALATSMLOACNGPGMKNKGOTGTLTGAGAGALLGSGFGKKGQL-VGVGVAL 63  
 QY 62 GGLIGSKIGSMDOQDK---IKLNSLEKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYN 117  
 D 64 GAVLGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEMRNDNGNYGYVTENKTY---- 119  
 QY 118 KOERROQYCREFOOKAMIAQOKOETIGTACPOPDGRMOVIS 158  
 D 120 -RNSTGQYCRETYOTYVIGSKQOKAIGNACROPDQOMQVYN 159

RESULT 4  
 G97860  
 17k surface antigen precursor [Imported] - Rickettsia conorii (strain Malish 7)  
 C:Species: Rickettsia conorii  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
 C:Accession: G97860  
 R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
 Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A:Reference number: A97700; MUID:21442074; PMID:11557893  
 A:Accession: G97860  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <KUR>  
 A:Cross-references: GB:AE006914; PIDN:AA030825.1; PID:915620425; GSPDB:GN00173  
 C:Accession: G97860  
 A:gene: omp  
 C:Superfamily: rickettsial common antigen

Query Match 34.3%; Score 285; DB 2; Length 159;  
 Best Local Similarity 37.9%; Pred. No. 3.1e-18;  
 Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;  
 QY 8 SLLIISV---FLVGC--AQNFSROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVL 61  
 D 5 SKIMIALATSMLOACNGPGMKNKGOTGTLTGAGAGALLGSGFGKKGQL-VGVGVAL 63  
 QY 62 GGLIGSKIGSMDOQDK---IKLNSLEKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYN 117  
 D 64 GAVLGQIGAGMDEQDRRLAELTSQRALETAPSGSNVEMRNDNGNYGYVTENKTY---- 119  
 QY 118 KOERROQYCREFOOKAMIAQOKOETIGTACPOPDGRMOVIS 158  
 D 120 -RNSTGQYCRETYOTYVIGSKQOKAIGNACROPDQOMQVYN 159

RESULT 5  
 C33971  
 Rickettsial common antigen precursor - Rickettsia typhi  
 C:Species: Rickettsia typhi  
 C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999  
 C:Accession: C33971  
 R:Anderson, B.E.; Tzianabos, T.  
 J. Bacteriol. 171, 5199-5201, 1989  
 A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.  
 A:Reference number: A33971; MUID:89359171  
 A:Accession: C33971  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <AND>  
 A:Cross-references: GB:M28481; NID:9152459; PIDN:AAA26377.1; PID:9152460  
 C:Superfamily: rickettsial common antigen

Query Match 33.2%; Score 276.5; DB 2; Length 159;  
 Best Local Similarity 39.6%; Pred. No. 1.8e-17;  
 Matches 55; Conservative 27; Mismatches 46; Indels 11; Gaps 4;  
 QY 25 SROEVGAATGAVVGAGVAGQLFGKSGRVSMAIG-GAVLGGLIGSKIGSMDOQDK---I 79  
 D 27 NKQGTGTLGAGAGALLGSGFGKKGQL-VGVGVALGAVLGQIGAGMDEQDRRLAEL 85  
 QY 80 KLNQSLKVKAGQVTRRNPDGTGNSYSVEPVRTYQRYNKOERROQYCREFOOKAMIAQOK 139  
 D 86 TSOALESPASGSNLEWRNDNGNYGYVTENKTY----RNSTGQYCRETYOTYVIGSKQ 140  
 QY 140 QETIGTACPOPDGRMOVIS 158  
 D 141 QETYGACROPDQOMQVYN 159

RESULT 6  
 A25972  
 17k antigen precursor - Rickettsia rickettsii  
 C:Species: Rickettsia rickettsii  
 C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 20-Aug-1999  
 C:Accession: A25972  
 R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; McDade, J.E.; Fu, Z.Y.  
 J. Bacteriol. 169, 2385-2390, 1987  
 A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia ricketts  
 A:Reference number: A25972; MUID:87222152  
 A:Accession: A25972

A:Molecule type: DNA  
A:Residues: 1-159 <AND>  
A:Cross-references: GB:M16486; NID:g152467; PIDN:AA26381.1; PID:g152468  
C:superfamily: rickettsial common antigen

Query Match 33.2%; Score 276; DB 2; Length 159;  
Best Local Similarity 37.3%; Pred. No. 2e-17;  
Matches 60; Conservative 31; Mismatches 54; Indels 16; Gaps 6;  
QY 8 SLLIIISV---FLVGC--AONFSROEVGAATGAVGVAGQLFGKSGRYSMAIG-GAVL 61  
D 5 SKLIIIALATSMIAQACNPGCMNKGITGLTGAGAGALLGSGFGKRGQL-VGVGVGALL 63  
QY 62 GGLIGSGISGSMDOOK-----IKLNSLEKVRKAGQVTRMRNPPTGNSYSVEPYRTYQRVN 117  
D 64 GAVLGQIGAGMDEQDRRLAELTSGRALETPSGSNVEMRNPNNGNYGYTPPKTY----- 119  
QY 118 KQERRQOYCRFQOKAMIAQKQEIYGTACPPDGRMOVIS 158  
D 120 -RNSTGQYCREYTGTVVIGGKQOKAYGACRQPDQMOVYN 159

RESULT 7  
D83169  
conserved hypothetical protein PA3819 [Imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83169  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MID:20437337  
A:Accession: D83169  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-182 <STO>  
A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AA607206.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3819

Query Match 15.3%; Score 127.5; DB 2; Length 182;  
Best Local Similarity 34.8%; Pred. No. 0.00037;  
Matches 32; Conservative 13; Mismatches 28; Indels 19; Gaps 2;  
QY 27 QEVGAATGAVGVAGQLFGKSGRYSMAIGAVLGGLIGSKIGSMDQDKI----- 79  
D 70 QINGTAIGAVVGGLLNQIGGCTGKATATVAGAVGGYAGNKVQEGMERDVTTETRC 129  
QY 80 -KLNOSLEK-----KAGQVTRMRNP 99  
D 130 STVHDSSEKVVGYDVKYMLDGKAGQIRMERDP 161

RESULT 8  
A13418  
17K surface antigen precursor [Imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: A13418  
R:Delvecchio, V.G.; Kaparatel, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens  
A:Reference number: AD3252; PMID:11756688  
A:Accession: A13418  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-131 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AA152516.1; PID:g17983328; GSPDB:GN00190  
A:Experimental source: strain 16M

C:Genetics:  
A:Gene: BME11335  
A:Map position: 1

Query Match 14.5%; Score 120.5; DB 2; Length 131;  
Best Local Similarity 31.7%; Pred. No. 0.0011;  
Matches 39; Conservative 17; Mismatches 44; Indels 23; Gaps 7;  
QY 46 GKSGGRYSMAIG-----AVLG-GLIGSKIGQ--SMDQDKIKLN-QSLEKAKG 91  
D 14 GKSGGPPS--LGSSSQKPEPTNLLASLNGLFNSASQLSAADRKKLEAEYRALEYSPAG 71  
QY 92 QVTRMRNPOTGNSYSVEPYRTYQRVKNKQERRQOYCRFQOKAMIAQKQEIYGTACPPD 151  
D 72 KSVLMSGAGS-NAGDVTAAQPIQ-----VGSQNCROYSHSFTITGGDOQVIRGTACRNP 124  
QY 152 GRW 154  
D 125 GSW 127

RESULT 9  
S23787  
outer membrane lipoprotein precursor - Yersinia enterocolitica  
C:Species: Yersinia enterocolitica  
C>Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 29-Sep-1999  
C:Accession: S23787  
R:Baumler, A.J.; Hantke, K.  
J. Bacteriol. 174, 1029-1035, 1992  
A:Title: A lipoprotein of Yersinia enterocolitica facilitates ferrioxamine uptake in  
A:Reference number: S23786; MID:92121089  
A:Accession: S23787  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-155 <BAE>  
A:Cross-references: EMBL:X60448; NID:g48577; PIDN:CAA42977.1; PID:g48579  
C:Superfamily: PAL cross-reacting lipoprotein

Query Match 13.5%; Score 112.5; DB 2; Length 155;  
Best Local Similarity 24.7%; Pred. No. 0.0068;  
Matches 37; Conservative 19; Mismatches 43; Indels 51; Gaps 4;  
QY 9 SLLIIISVFLVGAON-----FSROE-----VGAAT 33  
D 7 AVAIIAAVTLTGCAANNNTLSGDVFSASQAKOYQVTTGTLISVPVITIGGDDNNWGAIG 66  
QY 34 GAVVGVAGQLFGKSGRYSMAIGAVLGGLIGSKIGSMDQDKIKL----- 81  
D 67 GAVLGFLGNTVGGTGRSLATAAGAVAGGACGAGANNRTDGVQLEVRKDDGTTILV 126  
QY 82 --NQSLEKAKQVTRMRNPOTGNSYSVEP 109  
D 127 VOKQGPTRFSVGO--RVMLASGSGTYTVSP 154

RESULT 10  
B83514  
conserved hypothetical protein PA1053 [Imported] - Pseudomonas aeruginosa (strain PAO  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83514  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
radman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MID:20437337  
A:Accession: B83514  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-154 <STO>  
A:Cross-references: GB:AE004537; GB:AE004091; NID:g99446960; PIDN:AA604442.1; GSPDB:GN  
A:Experimental source: strain PA01





**THIS PAGE BLANK (USPTO)**





Db 491 GGAGAAAAAAGAGGEGIRGGAGGGGGLGGSGGSGRGLGGGAGAAAAAAGAGGGG 550  
Qy 64 LIGSKIGQ 71  
Db 551 LGGGGAGQ 558

## RESULT 2

US-09-247-806-1  
; Sequence 1, Application US/09247806  
; Patent No. 6280747  
; GENERAL INFORMATION:  
; APPLICANT: PHILIPPE, Michel  
; APPLICANT: GARSON, Jean-Claude  
; APPLICANT: ARRAUDEAU, Jean-Pierre  
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT  
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN  
; FILE REFERENCE: 6388-0365-0  
; CURRENT FILING DATE: 1999-02-11  
; EARLIER APPLICATION NUMBER: US/09/247,806  
; EARLIER FILING DATE: 1998-02-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Nephila clavipes  
US-09-247-806-1

## Query Match

Best Local Similarity 9.4%; Score 78.5; DB 4; Length 651;  
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

Qy 27 QEVGATGAVVG---GVAGQ-----LFGKSGRVSMAIGAVLGG 63  
Db 491 GGAGAAAAAAGAGGEGIRGGAGGGGGLGGSGGSGRGLGGGAGAAAAAAGAGGGG 550  
Qy 64 LIGSKIGQ 71  
Db 551 LGGGGAGQ 558

## RESULT 3

US-08-425-069-2  
; Sequence 2, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; APPLICANT: Himman, Michael B.  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5728810th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/425,069  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-425-069-2

## Query Match

Best Local Similarity 9.4%; Score 78.5; DB 1; Length 718;  
Matches 24; Conservative 4; Mismatches 17; Indels 23; Gaps 2;

Qy 27 QEVGATGAVVG---GVAGQ-----LFGKSGRVSMAIGAVLGG 63  
Db 491 GGAGAAAAAAGAGGEGIRGGAGGGGGLGGSGGSGRGLGGGAGAAAAAAGAGGGG 550  
Qy 64 LIGSKIGQ 71  
Db 551 LGGGGAGQ 558

## RESULT 4

US-08-317-844B-2  
; Sequence 2, Application US/08317844B  
; Patent No. 5989894  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; APPLICANT: Himman, Michael B.  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5989894th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,844B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-105P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-317-844B-2

## Query Match

9.4%; Score 78.5; DB 2; Length 718;



```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2029060
US-09-067-351-2

Query Match
Best Local Similarity 9.3%; Score 77; DB 2; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 7 GSSLIIIVFLVGCANFSROEVGATGAVGVAGQLFGKSGRVSMAIGAVLGGLIG 66
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 54 GASFGSRSLYINLGAKRVLNCGSSCSRGFGGRASNGFGVNSG---FGYGGVGGGFSG 110
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 67 SK-----IGQMDQODKIKLNSLEKVKACQVTRMNPDTGNSYSVEPVRTY 113
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 111 PSFPVCPGIGQIEVTVNOSLITPLHLQIDPTIQRVAREERQIKTLNKFISFDKVRFL 170
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 114 QRYNK 118
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 171 EQQNK 175
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1

RESULT 8
US-09-360-490-2
; Sequence 2, Application US/09360490
; Patent No. 6221843
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah
; TITLE OF INVENTION: HUMAN KERATINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/360,490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/067,351
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0511 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
```

```

; CLONE: 2029060
US-09-360-490-2

Query Match
Best Local Similarity 9.3%; Score 77; DB 4; Length 551;
Matches 28; Conservative 28; Mismatches 53; Indels 16; Gaps 2;

QY 7 GSSLIIIVFLVGCANFSROEVGATGAVGVAGQLFGKSGRVSMAIGAVLGGLIG 66
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 54 GASFGSRSLYINLGAKRVLNCGSSCSRGFGGRASNGFGVNSG---FGYGGVGGGFSG 110
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 67 SK-----IGQMDQODKIKLNSLEKVKACQVTRMNPDTGNSYSVEPVRTY 113
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 111 PSFPVCPGIGQIEVTVNOSLITPLHLQIDPTIQRVAREERQIKTLNKFISFDKVRFL 170
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 114 QRYNK 118
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 171 EQQNK 175
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1

RESULT 9
US-08-374-077C-2
; Sequence 2, Application US/08374077C
; Patent No. 6027912
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha1
; TITLE OF INVENTION: Calcium Channel Subunit
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,077C
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MCGOWAN, MALCOLM M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-374-077C-2

Query Match
Best Local Similarity 9.2%; Score 76.5; DB 3; Length 2516;
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

QY 29 VQATGAVGVAGQLFGKSGRVSMAIGG--AVGGLIGSKIGS---MDQODKIKLN 82
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
Db 2389 ICGSSNGSIFGSGAGGLGAGSGGVG--GLGSSSINNAFGSGSGSSLSPOHOPISGLTN 2447
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
QY 83 -----QSLKVKACQVTRMNPDTGNSYSVEPVRTYQRYNKQERRQOYCREFOQAMIA 136
   1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1 1:1
```

DB 2448 SPIIPDNRRLRRVATVTTTNNNNKSOVSQNNSSLLVNRANANSONMNSPTGQVQOOSPLR 2507  
QY 137 GOKOEIYGT 145  
DB 2508 GOGNQTYS 2516

## RESULT 10

US-08-895-590-2  
; Sequence 2, Application US/08895590  
; Patent No. 6207410  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Linda M.  
; APPLICANT: Ren, Dejian  
; APPLICANT: Zheng, Wei  
; APPLICANT: Dubaig, Manuel Marcel Paul  
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel  
; NUMBER OF SEQUENCES: 101  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/895,590  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374,888  
; FILING DATE: 19-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm M.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 022650-263  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-895-590-2

Query Match 9.2%; Score 76.5; DB 4; Length 2516;  
Best Local Similarity 24.8%; Pred. No. 17;  
Matches 32; Conservative 18; Mismatches 66; Indels 13; Gaps 4;

QY 29 VCAATGAVVGVAGQLFGKSGRYNAIGG-AVLGLISKTIQS-----MDQODKIKLN 82  
DB 2389 IGSSNGSIFGSGAGLGAGSGGVG-GLGSSSIRNAFGSGSGSPSLSPQHQPYSGLTN 2447  
QY 83 -----QSLKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNKQERROQYCREFOQKAMIA 136  
DB 2448 SPIIPDNRRLRRVATVTTTNNNNKSOVSQNNSSLLVNRANANSONMNSPTGQVQOOSPLR 2507  
QY 137 GOKOEIYGT 145  
DB 2508 GOGNQTYS 2516

RESULT 11  
US-07-803-633A-13  
; Sequence 13, Application US/07803633A  
; Patent No. 5369025

; GENERAL INFORMATION:  
; APPLICANT: NAZERIAN, Keyvan  
; APPLICANT: LEE, Lucy F.  
; APPLICANT: YANAGIDA, No. 53690250ru  
; APPLICANT: OGAMA, Ryohel  
; APPLICANT: LI, YI  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VACCINE FOR  
; PROTECTION AGAINST MAREK'S DISEASE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 301 No. 5369025th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/803,633A  
; FILING DATE: 19911210  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy JF., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1644-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 865 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-803-633A-13

Query Match 9.1%; Score 76; DB 1; Length 865;  
Best Local Similarity 25.9%; Pred. No. 4.7;  
Matches 37; Conservative 30; Mismatches 46; Indels 30; Gaps 8;

QY 16 FLVCAQNFESRQ-EVGAATGAVVGVAGQLFGKSGRVS-----MAIGAVLGLI 65  
DB 688 FMNGLAEFLFMGQVGOAIGKVVVGAAGAIVSTISG-VSAFMSNPGALAIIGLIIAGLV 746  
QY 66 GSKIGOSMDQODKIKLN-----QSLKVKAGQVTRMRNPDTGNSYSVEPVRTYQRYNK 118  
DB 747 AAFI--AYRYVNRKLSPMKALYPMTEVLKA-QATRELHESDD-----LERTSI 795  
QY 119 QERRQCYCRE-FOOKAMIAQOKQ 140  
DB 796 DERKLEAREMITYMALVSAER 818

RESULT 12  
US-08-220-151-6  
; Sequence 6, Application US/08220151  
; Patent No. 5529780  
; GENERAL INFORMATION:  
; APPLICANT: Paoletti, Enzo  
; APPLICANT: Limbach, Keith J.  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
; CANINE HERPESVIRUS GB, GC AND GD AND USES THEREFOR  
; NUMBER OF SEQUENCES: 91  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY

COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/220,151  
FILING DATE: 30-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2540  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-220-151-6

Query Match 9.0%; Score 74.5; DB 1; Length 913;  
Best Local Similarity 23.6%; Pred. No. 7.4;  
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCQNF-SROEVGAATGAVGVAGQLFGKSGRVS-----MAIGAVLGGLI 65  
DB 753 VLLRGIANFQGLGDVGAAGVKVILGATGAVISAVGAWSFSLNPGALALGLVLVAGLV 812  
QY 66 GSKT-----GQSMDDQDKIKLNSLEKVK-----A 90  
DB 813 AAFLAYRHISRLRNPMKALYPTTKLEKEDVDEGDVDEKLDQARDMIRYSIVSALE 872  
QY 91 GQYTRRNPDPTGNSYSVEPY-----RTYQRYNKOE 120  
DB 873 QOEHKARKKNSGALLASRVGAMATRRRHRYQRLSESD 909

RESULT 13  
US-08-413-118-6  
Sequence 6, Application US/08413118  
Patent No. 5688920  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/413,118  
FILING DATE: 29-MAR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/220,151  
FILING DATE: 30-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-413-118-6

Query Match 9.0%; Score 74.5; DB 1; Length 913;  
Best Local Similarity 23.6%; Pred. No. 7.4;  
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCQNF-SROEVGAATGAVGVAGQLFGKSGRVS-----MAIGAVLGGLI 65  
DB 753 VLLRGIANFQGLGDVGAAGVKVILGATGAVISAVGAWSFSLNPGALALGLVLVAGLV 812  
QY 66 GSKT-----GQSMDDQDKIKLNSLEKVK-----A 90  
DB 813 AAFLAYRHISRLRNPMKALYPTTKLEKEDVDEGDVDEKLDQARDMIRYSIVSALE 872  
QY 91 GQYTRRNPDPTGNSYSVEPY-----RTYQRYNKOE 120  
DB 873 QOEHKARKKNSGALLASRVGAMATRRRHRYQRLSESD 909

RESULT 14  
US-08-473-446-6  
Sequence 6, Application US/08473446  
Patent No. 6017542  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,446  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/413,118  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-473-446-6

Query Match 9.08; Score 74.5; DB 3; Length 913;  
Best Local Similarity 23.68; Pred. No. 7.4;  
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCANF-SROEVGAATGAVGVAGOLFSGSGRVS-----MAIGAVLGGLI 65  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
DB 753 VLLRGIANFQGLDVGAAVGVVVGATGAVISAVGKVSFLSNPFGALATGLVLAGLV 812  
QY 66 GSKI-----GQSMDOODKIKLNQSLKVK-----A 90  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
DB 813 AAFLAYRHISRLRRNPKKALPYTTKTLEKGVDEGVDEAKLDQARDMIRMSIVSALE 872  
QY 91 GQVTRWRNPDTGNSYSVEPV-----RTYQRYNKOE 120  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
DB 873 QQEHKARKKNSGALLASRVGAMATRRRHYQRLSESD 909

RESULT 15  
US-09-232-468A-2  
; Sequence 2, Application US/09232468A  
; Patent No. 6207165  
; GENERAL INFORMATION:  
; APPLICANT: AUDONNET et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE  
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES  
; FILE REFERENCE: 454313-2230  
; CURRENT APPLICATION NUMBER: US/09/732.468A  
; CURRENT FILING DATE: 1999-01-05  
; NUMBER OF SEQ. ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 913  
; TYPE: PRT  
; ORGANISM: Pseudorabies virus  
US-09-232-468A-2

Query Match 9.08; Score 74.5; DB 4; Length 913;  
Best Local Similarity 23.68; Pred. No. 7.4;  
Matches 37; Conservative 18; Mismatches 51; Indels 51; Gaps 5;

QY 15 VFLVGCANF-SROEVGAATGAVGVAGOLFSGSGRVS-----MAIGAVLGGLI 65  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
DB 753 VLLRGIANFQGLDVGAAVGVVVGATGAVISAVGKVSFLSNPFGALATGLVLAGLV 812  
QY 66 GSKI-----GQSMDOODKIKLNQSLKVK-----A 90  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
DB 813 AAFLAYRHISRLRRNPKKALPYTTKTLEKGVDEGVDEAKLDQARDMIRMSIVSALE 872  
QY 91 GQVTRWRNPDTGNSYSVEPV-----RTYQRYNKOE 120  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
DB 873 QQEHKARKKNSGALLASRVGAMATRRRHYQRLSESD 909

Search completed: October 27, 2002, 11:12:13  
Job time : 9.34197 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:58:51 : Search time 21.4111 Seconds  
(without alignments)  
835.217 Million cell updates/sec

Title: US-09-677-374-4  
Perfect score: 832  
Sequence: 1 MRGCLQGSSLLIISVFLVGC.....ITGTACPPDGRMQVISTFEK 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_032802:\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	832	100.0	161	22	AA881127
2	832	100.0	256	22	AA881128
3	815	98.0	162	22	AA881126
4	815	98.0	162	22	AA881126
5	112	13.5	20	22	AA881130
6	111	13.3	224	22	AA820105
7	102.5	12.3	223	20	AA134487
8	102.5	12.3	230	20	AA134362
9	101	12.1	154	11	AA805799
10	95	11.4	309	22	AA815906
11	88	10.6	528	22	AA882611

12	87.5	10.5	147	22	AB869847
13	86.5	10.4	2309	22	AB866232
14	84	10.1	1251	22	AB861254
15	83	10.0	116	19	AA111028
16	83	10.0	423	22	AB30695
17	82	9.9	2017	22	AB806301
18	81	9.7	666	22	AB858019
19	80.5	9.7	542	22	AB85790
20	80.5	9.7	542	22	AB85791
21	80.5	9.7	542	22	AB870501
22	79.5	9.6	618	21	AB856803
23	78.5	9.4	651	20	AA140097
24	78.5	9.4	718	12	AA14308
25	78.5	9.4	718	19	AA853346
26	78.5	9.4	718	21	AA59070
27	77.5	9.3	102	22	AA41943
28	77.5	9.3	255	21	AA811398
29	77.5	9.3	255	22	AA874199
30	77.5	9.3	255	22	AA870769
31	77.5	9.3	255	22	AA772020
32	77.5	9.3	285	20	AA41688
33	77.5	9.3	285	21	AB844244
34	77.5	9.3	285	22	AA029025
35	77.5	9.3	285	22	AA839011
36	77.5	9.3	302	22	AA40157
37	77.5	9.3	354	22	AA19445
38	77.5	9.3	514	22	AA36520
39	77	9.3	223	21	AA857121
40	77	9.3	285	21	AA873440
41	77	9.3	285	21	AA873440
42	77	9.3	285	22	AA853288
43	77	9.3	551	21	AA52398
44	77	9.3	2599	21	AA75098
45	76.5	9.2	2516	17	AAW01875

#### ALIGNMENTS

RESULT 1	
AA881127	AA881127 standard; Protein: 161 AA.
XX	XX
AC	AA881127:
XX	XX
DT	11-JUL-2001 (first entry)
XX	XX
DE	Optimised Ospa protein 17E2 amino acid sequence.
XX	XX
KW	Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW	vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
KW	SRS.
XX	XX
OS	Piscirickettsia salmonis.
OS	Synthetic.
XX	XX
FH	Key
FT	Region
FT	109..128
FT	/label= B_cell_epitope
PN	CA2281913-A1.
XX	XX
PD	17-MAR-2001.
XX	XX
XX	17-SEP-1999; 99CA-2281913.
PF	17-SEP-1999; 99CA-2281913.
XX	XX
PR	17-SEP-1999; 99CA-2281913.
XX	XX
PA	(KAYW/) KAY W W.
PA	(BURY/) BURIAN J.
PA	(KUZV/) KUZV M A.
XX	XX
PI	Kay WW, Burian J, Kuzv MA;

Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
H. pylori ORF 01cp  
A fusion of anti-C  
Novel human diagno  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Human prostate can  
Spider silk protei  
N. clavipes draglin  
Nephila clavipes s  
N. clavipes spider  
Human polyptide  
E. coli expression  
pe1B-scfoxa2olon  
Expression plasmid  
E. carotovora pe1B  
Human PRO284 (UNO2  
Human PRO polypt  
Human polyptide  
Human polyptide  
Human diagnostic a  
Pseudomonas aerugi  
Human prostate can  
Human signal pepti  
Human secreted pro  
Human protein HP10  
Human keratin KERT  
Neisseria meningit  
Neuronal Invertebr

```

XX DR WPI: 2001-316844/34.
XX DR N-PSDB: AAF86247.
XX PT Method for protecting polkilohermic fish against salmonid rickettsial
XX PT septicemia and other rickettsial diseases comprises administering a
XX PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX PS
XX PS Disclosure: Fig 5; 35pp; English.
XX CC This invention relates to a method for the protection against infection
XX CC of a polkilohermic fish by the bacterial pathogen, Piscirickettsia
XX CC salmonis. The method comprises administering an immunogenic amount of a
XX CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
XX CC Ospa in the form of a vaccine. The method is used for protecting animals,
XX CC particularly polkilohermic fish, against the bacterial pathogen
XX CC P. salmonis. The method is also useful for protecting against salmonid
XX CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX CC sequence represents optimised P. salmonis Ospa protein 17E2. The DNA
XX CC encoding Ospa 17E2 (AAF86247) has been optimised for expression in
XX CC Escherichia coli. An Ospa protein with an N-terminal fusion partner is
XX CC used in a vaccine to create an anti-Ospa antibody response.
XX SQ Sequence 161 AA:
XX
XX Query Match 100.0%; Score 832; DB 22; Length 161;
XX Best Local Similarity 100.0%; Pred. No. 3e-81;
XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MRGCLGSSLIITISVFLVGCACNFSROEVGATGAVGVAGQLFGKSGRVSMAIGAV 60
XX DB 1 MRGCLGSSLIITISVFLVGCACNFSROEVGATGAVGVAGQLFGKSGRVSMAIGAV 60
XX
XX QY 61 LGLIGSKTIGQSMDOODKIKLNQSLKVKACGVTRMRNPDGNSYSVPVTRYRKNQE 120
XX DB 61 LGLIGSKTIGQSMDOODKIKLNQSLKVKACGVTRMRNPDGNSYSVPVTRYRKNQE 120
XX
XX QY 121 RROQYCREFOQKAMTAGOKOEITGTACPOPDRGROWISTEK 161
XX DB 121 RROQYCREFOQKAMTAGOKOEITGTACPOPDRGROWISTEK 161
XX
XX RESULT 2
XX AAB81128
XX ID AAB81128 standard; Protein: 256 AA.
XX AC AAB81128;
XX DT 11-JUL-2001 (first entry)
XX DE C17E2 Ospa construct with N-terminal fusion partner.
XX KW Polkilohermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;
XX KW SRS; 17E2; fusion construct.
XX OS Piscirickettsia salmonis.
XX OS Synthetic.
XX FH Key
XX FH Region 1..95 Location/Qualifiers
XX FT /label= Undefined; N-terminal_fusion_partner
XX FT /label= C17E2_Ospa
XX FT /note= "Product of Ospa gene optimised for expression in
XX FT Escherichia coli"
XX CA2281913-A1.
XX PN
XX PD 17-MAR-2001.
XX PF 17-SEP-1999; 99CA-2281913.
XX

```

```

PR 17-SEP-1999; 99CA-2281913.
XX
XX PA (KAYW/) KAY W W.
XX PA (BURI/) BURIAN J.
XX PA (KUZY/) KUZYSK M A.
XX
XX PI Kay WW, Burian J, Kuzysk MA;
XX
XX DR WPI: 2001-316844/34.
XX DR N-PSDB: AAF86248.
XX
XX PT Method for protecting polkilohermic fish against salmonid rickettsial
XX PT septicemia and other rickettsial diseases comprises administering a
XX PT vaccine containing the Ospa protein of Piscirickettsia salmonis
XX PS
XX PS Example 4; Fig 5; 35pp; English.
XX
XX CC This invention relates to a method for the protection against infection
XX CC of a polkilohermic fish by the bacterial pathogen, Piscirickettsia
XX CC salmonis. The method comprises administering an immunogenic amount of a
XX CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a
XX CC Ospa in the form of a vaccine. The method is used for protecting animals,
XX CC particularly polkilohermic fish, against the bacterial pathogen
XX CC P. salmonis. The method is also useful for protecting against salmonid
XX CC rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX CC sequence represents the amino acid sequence of C17E2, a P. salmonis Ospa
XX CC construct optimised for expression in Escherichia coli, fused to an
XX CC undefined N-terminal fusion partner. The fusion protein is used in a
XX CC vaccine to create an anti-Ospa antibody response.
XX SQ Sequence 256 AA:
XX
XX Query Match 100.0%; Score 832; DB 22; Length 256;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-81;
XX Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MRGCLGSSLIITISVFLVGCACNFSROEVGATGAVGVAGQLFGKSGRVSMAIGAV 60
XX DB 96 MRGCLGSSLIITISVFLVGCACNFSROEVGATGAVGVAGQLFGKSGRVSMAIGAV 155
XX
XX QY 61 LGLIGSKTIGQSMDOODKIKLNQSLKVKACGVTRMRNPDGNSYSVPVTRYRKNQE 120
XX DB 156 LGLIGSKTIGQSMDOODKIKLNQSLKVKACGVTRMRNPDGNSYSVPVTRYRKNQE 215
XX
XX QY 121 RROQYCREFOQKAMTAGOKOEITGTACPOPDRGROWISTEK 161
XX DB 216 RROQYCREFOQKAMTAGOKOEITGTACPOPDRGROWISTEK 256
XX
XX RESULT 3
XX AAG78025
XX ID AAG78025 standard; Protein: 162 AA.
XX AC AAG78025;
XX DT 15-JAN-2002 (first entry)
XX DE Piscirickettsia salmonis polypeptide p10.6.
XX KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
XX KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;
XX KW ATCC VR-1361.
XX OS Piscirickettsia salmonis.
XX OS
XX PN WO200168865-A2.
XX PD 20-SEP-2001.
XX PF 12-MAR-2001; 2001WO-GB01055.
XX PR 11-MAR-2000; 2000GB-0005838.
XX PR 01-JUL-2000; 2000GB-0016080.
XX

```

```

PR 01-JUL-2000; 2000GB-0016082.
PR 29-JUL-2000; 2000GB-0018599.
XX
XX (AQUA-) AQUA HEALTH EURO LTD.
XX
XX Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
PI Burzio L;
XX
XX WPI: 2001-639050/73.
DR N-PSDB; AAF79040.
XX
XX New nucleic acids encoding an amino acid sequence homologous to the
PT surface antigen present on Piscirickettsia salmonis are useful to
PT protect fish against piscirickettsiosis.
XX
XX Claim 6: Fig 5; 25pp: English.
XX
XX The invention relates to nucleic acid sequences and the encoded protein
CC of a least part of the surface antigen present on Piscirickettsia
CC salmonis for production of a vaccine with antibacterial activity to
CC protect fish against P. salmonis which causes piscirickettsiosis, also
CC known as salmonid rickettsial septicaemia.
XX
XX Sequence 162 AA:
SQ
Query Match 98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 2 RGLQSSLLIIISVFLVGCQNFSROEVGAATGAVVGVAGOLFQKSGSVMAIGAVL 61
DB 3 RGLQSSLLIIISVFLVGCQNFSROEVGAATGAVVGVAGOLFQKSGSVMAIGAVL 62
OY 62 GGLIGSKIGSMDQDKIKLNOSLEKVKAGQVTRMRNPDGNSYSVEPVRTYGRYNKQER 121
DB 63 GGLIGSKIGSMDQDKIKLNOSLEKVKAGQVTRMRNPDGNSYSVEPVRTYGRYNKQER 122
OY 122 ROOYCREFOOKAMIAQOKOIYGTACPDGPRMOWISTEK 161
DB 123 ROOYCREFOOKAMIAQOKOIYGTACPDGPRMOWISTEK 162
RESULT 4
AAB81126
ID AAB81126 standard; Protein; 162 AA.
XX
XX AAB81126;
AC
XX
XX 11-JUL-2001 (first entry)
DT
XX
XX Ospp antigen amino acid sequence.
DE
XX
XX polkilohermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospp; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS.
XX
XX Piscirickettsia salmonis.
OS
XX
XX Key Location/Qualifiers
FH 110..129
FT Region /label= B_cell_epitope
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAWW/) KAY W W.
XX (BURJ/) BURIAN J.
XX (KUZV/) KUZYK M A.

```

```

XX
XX Kay WW, Burian J, Kuzyk MA;
PI
XX
XX WPI: 2001-316844/34.
DR N-PSDB; AAF86246.
XX
XX Method for protecting polkilohermic fish against salmonid rickettsial
PT septicaemia and other rickettsial diseases comprises administering a
PT vaccine containing the Ospp protein of Piscirickettsia salmonis.
XX
XX Example 2: Fig 2B; 35pp: English.
XX
XX This invention relates to a method for the protection against infection
CC of a polkilohermic fish by the bacterial pathogen, piscirickettsia
CC salmonis. The method comprises administering an immunogenic amount of a
CC P. salmonis specific antigen termed Ospp, or an immunogenic fragment of
CC Ospp in the form of a vaccine. The method is used for protecting animals,
CC particularly polkilohermic fish, against the bacterial pathogen
CC P. salmonis. The method is also useful for protecting against salmonid
CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present
CC sequence represents P. salmonis Ospp protein. An Ospp protein with an
CC N-terminal fusion partner is used in a vaccine to create an anti-Ospp
CC antibody response.
XX
XX Sequence 162 AA:
SQ
Query Match 98.0%; Score 815; DB 22; Length 162;
Best Local Similarity 98.8%; Pred. No. 2e-79;
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 2 RGLQSSLLIIISVFLVGCQNFSROEVGAATGAVVGVAGOLFQKSGSVMAIGAVL 61
DB 3 RGLQSSLLIIISVFLVGCQNFSROEVGAATGAVVGVAGOLFQKSGSVMAIGAVL 62
OY 62 GGLIGSKIGSMDQDKIKLNOSLEKVKAGQVTRMRNPDGNSYSVEPVRTYGRYNKQER 121
DB 63 GGLIGSKIGSMDQDKIKLNOSLEKVKAGQVTRMRNPDGNSYSVEPVRTYGRYNKQER 122
OY 122 ROOYCREFOOKAMIAQOKOIYGTACPDGPRMOWISTEK 161
DB 123 ROOYCREFOOKAMIAQOKOIYGTACPDGPRMOWISTEK 162
RESULT 5
AAB81130
ID AAB81130 standard; Peptide; 20 AA.
XX
XX AAB81130;
AC
XX
XX 11-JUL-2001 (first entry)
DT
XX
XX Ospp B-cell epitope peptide #2.
DE
XX
XX polkilohermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; Ospp; salmonid rickettsial septicaemia; rickettsial disease;
KW SRS; antibody.
XX
XX Piscirickettsia salmonis.
OS
XX
XX Key Location/Qualifiers
FH 110..129
FT Region /label= B_cell_epitope
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAWW/) KAY W W.
XX (BURJ/) BURIAN J.
XX (KUZV/) KUZYK M A.
XX
XX Kay WW, Burian J, Kuzyk MA;
PI
XX

```

DR WPI: 2001-316844/34.

XX Method for protecting poikilothermic fish against salmonid rickettsial  
PT septicemia and other rickettsial diseases comprises administering a  
PT vaccine containing the OspA protein of *Piscirickettsia salmonis*.  
XX  
XX Example 2: Page 17, 35pp; English.

CC This invention relates to a method for the protection against infection  
CC of a poikilothermic fish by the bacterial pathogen, *Piscirickettsia*  
CC salmonis. The method comprises administering an immunogenic amount of a  
CC *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of a  
CC OspA in the form of a vaccine. The method is used for protecting animals,  
CC particularly poikilothermic fish, against the bacterial pathogen  
CC *P. salmonis*. The method is also useful for protecting against salmonid  
CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
CC sequence represents an immunogenic epitope of the *P. salmonis* OspA  
XX protein. The peptide is used to raise rabbit anti-OspA antibodies.  
SQ Sequence 20 AA;

Query Match 13.5%; Score 112; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.3e-05;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 109 PVRTYGRYKQRRQCYCRE 128  
DB 1 PVRTYGRYKQRRQCYCRE 20

RESULT 6

AAB20105  
ID AAB20105 standard; Protein: 224 AA.

XX AAB20105;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB13 protein.

KW BASB13: infection; otitis media; pneumonia; therapy; diagnosis;  
KW antibacterial; antimicrobial.

OS Moraxella catarrhalis.

XX WO200100836-A1.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-EP05851.

PR 25-JUN-1999; 99GB-0015044.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonard J;

DR WPI: 2001-112458/12.

DR N-PSDB; AAF30043.

XX New BASB13 polypeptide isolated from Moraxella catarrhalis bacterium,  
PT useful for diagnosing and producing vaccines against bacterial  
PT infections such as otitis media and pneumonia  
XX  
XX Claim 1: Page 67; 86pp; English.

CC The present sequence is that of BASB13 protein from Moraxella  
CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis  
CC media in children and pneumonia in adults. The invention provides  
CC BASB13 polypeptides, and polynucleotides encoding them, as well as  
CC expression vectors, host cells and methods for producing BASB13  
CC polypeptides using recombinant methods. Also claimed is a vaccine  
CC composition comprising a BASB13 polypeptide, an immunogenic

CC fragment of a BASB13 polypeptide, or a polypeptide having at least  
CC 85% amino acid sequence identity to BASB13, or comprising a  
CC polynucleotide encoding such a polypeptide. A claimed method of  
CC diagnosing a Moraxella infection involves identifying a BASB13  
CC polypeptide or antibody. A claimed therapeutic composition useful  
CC in treating humans with M. catarrhalis infection comprises at least  
CC 1 antibody directed against a BASB13 polypeptide. BASB13  
CC polypeptides also have utility in raising specific antibodies,  
CC and in screening for antibacterial drugs.

SQ Sequence 224 AA;

Query Match 13.3%; Score 111; DB 22; Length 224;  
Best Local Similarity 33.3%; Pred. No. 0.00097;  
Matches 37; Conservative 15; Mismatches 35; Indels 24; Gaps 5;

OY 7 GSSLIIISVFLVGCANQNSRQEVGATG-----AVVGVAGQLFGSGRVSMAITG 58  
DB 7 GVLLASMSALAGCANNGT---TGNGTFCGANVKNRAVIGAVAGAL---GCTAISKATGG 60  
OY 59 -----AVLGGLIGSKTIGQSMDOODKIKLQSLSEKVKAGQVTRMKNPTGN 103  
DB 61 EKTGRDAILGAAGAAGAAAGAYMERQAK---QIEQCMQGTGTVTHDTPGN 107

RESULT 7

AAV34487  
ID AAV34487 standard; Protein: 223 AA.

XX AAV34487;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG3.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
KW vaccine; antigenic.

OS Porphyromonas gingivalis.

XX WO9929670-A1.

XX 17-JUN-1999.

PD 10-DEC-1998; 98WO-AU01023.

PF 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSTC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
PI Ross BC, Rothel LJ, Webb EA.

DR WPI: 1999-385613/32.

DR N-PSDB; AAX91705.

XX Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis  
XX

PS Claim 1: Page 469; 588pp; English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX91318 to  
CC AAX91583. AAX91802 to AAX91989 represent PCR primers used in the

used as vaccines especially against Borborymonas ovinale probes can

Query Match	12.1%;	Score 101;	DB 11;	Length 154;
Best Local Similarity	39.6%;	Pred. No. 0.0071;		
Matches	21;	Conservative	9;	Mismatches 23;
				Indels 0;
				Gaps 0

*[The page contains faint, illegible markings or bleed-through from the reverse side.]*

Db 62 VGTGGGALGGIAGSTIGRGQALAAVYGALGALAGSKIEKMSQVNAEL 114

## RESULT 10

ABG15906  
ID ABG15906 standard; Protein; 309 AA.  
XX  
AC ABG15906;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15897.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
XX N-PSDB: AAS80093.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT biologically for genetic disorders or other traits and to assess  
PT biodiversity

Claim 20; SEQ ID No 46265; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping; and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations in  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 309 AA;

Query Match 11.4%; Score 95; DB 22; Length 309;  
Best Local Similarity 33.0%; Pred. No. 0.077;

Matches 32; Conservative 12; Mismatches 39; Indels 14; Gaps 3;

QY 12 IISVFLVCAQNFNR-----QEVGATGAVGVGVGQJFGNG--SGRVSMAIGAVLIGL 64  
DB 185 IISILPAKVAVDNSQKRNKAQAFGLICAVAGVIGHWVSGSSMTTACAGVAGVACAA 244

QY 65 IGSKI-----GQSMQDQDKIKLNQSLSEKVRAGQVT 94  
DB 245 AGSMVNDKTIMEGVSLTYKEGTRKYVTSTQVGKEROFT 281

## RESULT 11

AAB82611  
ID AAB82611 standard; Protein; 528 AA.  
XX  
AC AAB82611;  
XX  
DT 02-OCT-2001 (first entry)  
XX  
DE Spider recombinant silk protein PETNCDs.  
XX  
KW Spider; orb-weaver; silk protein; PETNCDs; structural protein;  
XX purification; fibre; spinning.  
XX  
OS Nephila clavipes.  
XX  
FH Key Location/Qualifiers  
FH  
FT Misc-difference 417  
FT /note= "encoded by TAT"  
FT Misc-difference 427  
FT /note= "encoded by CGA"  
FT Misc-difference 522  
FT /note= "encoded by GAG"  
XX  
PN WO200153333-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 01-NOV-2000; 2000WO-US30086.  
XX  
PR 20-JAN-2000; 2000US-0490291.  
XX  
PA (MELLO/) MELLO C M.  
PA (ARCTI/) ARCIDIAONO S.  
PA (BUTL/) BUTLER M M.  
PA (USSA ) US SEC OF ARMY.  
XX  
PI Mello CM, Arcidiacono S, Butler MM;  
XX  
DR WPI: 2001-483136/52.  
XX N-PSDB: AAH26304.  
XX

PT Recovering structural polypeptides in a biological sample, useful for  
PT purifying and spinning spider silks and other structural proteins,  
PT comprises treating the sample containing the polypeptides with an acid

Claim 2; Page 41-42; 49pp; English.

XX The present sequence is that of the orb-weaver spider (Nephila  
XX clavipes) recombinant silk protein PETNCDs. The invention  
XX provides methods for purifying and spinning spider silks and other  
XX structural proteins. Organic acids are used to lyse recombinant  
XX cells or other biological samples (such as non-recombinantly  
XX derived cells), and enrich the purity and yields of structural  
XX proteins by hydrolysing many of the macromolecules while leaving  
XX the structural proteins intact. In the case of silk proteins, the  
XX resulting lysate is further purified by ion-exchange or affinity  
XX chromatography and processed into an aqueous-based mixture for  
XX fibre spinning. In the present case, the PETNCDs gene was cloned  
XX into vector pET24 for recombinant expression in Escherichia coli,  
XX and recombinant silk protein was obtained using formic acid  
XX containing denaturant (3 M guanidine-HCl) and by affinity  
XX chromatography on nickel-NTA agarose. Products obtained using the  
XX methods of the invention can be used in the construction of many  
XX materials including films, fibres, woven articles, sutures,  
XX ballistic protection, parachutes and parachute cords. The new  
XX method has the following advantages over prior art: it involves  
XX fewer steps, requires less time and smaller volumes of reagents,

CC results in better recovery of protein at higher purity (70-99%),  
CC is easy to scale up, and the fibres are spun in an environmentally  
CC benign solution reducing hazardous waste accumulation and cost.

XX Sequence 528 AA;

Query Match 10.6%; Score 88; DB 22; Length 528;

Best Local Similarity 48.9%; Pred. No. 0.87; Mismatches 19; Indels 2; Gaps 1;

Matches 23; Conservative 3; Mismatches 19; Indels 2; Gaps 1;

Qy 27 QEVGATGAVGVAGQ-LFGKSGRVSMAIGAVLGKIGSKIGQ 71

Db 320 QGAGAAAAAGAGAGGCGGAGCAAAAAAGAGAGCGGCGAGQ 366

RESULT 12

ABB69847 standard; Protein; 147 AA.

XX ABB69847;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 36333.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN MO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB; ABLI3950.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Disclosure; SEQ ID NO 36333; 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 147 AA;

Query Match 10.5%; Score 87.5; DB 22; Length 147;

Best Local Similarity 27.5%; Pred. No. 0.19; Mismatches 39; Gaps 7;

Matches 36; Conservative 18; Mismatches 38; Indels 39; Gaps 7;

Qy 36 VVGAVQALFGKSGRVSMA-----IGAVLGLIGSKIGQSDQ---DKIKLN 82

Db 11 VTSGISS--IGLGNRRFSLSNQSSSQSGIGAIPEGIVSGAQQQQQVPPQPLQ 68

Qy 83 OSLEKVKAGQVTRMRNPDGNSYSVEPVTRYORYNKQEROQYCR-EFOOKAMIAQKOE 141

Db 69 QSLPQQQQQ-----00000001000HGHAPLQ000AST-GSGLG 106

Qy 142 IYGTACPPDG 152

Db 107 YGTAPR-PTG 116

RESULT 13

ABB66232 standard; Protein; 2309 AA.

XX ABB66232;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 25488.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN MO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB; ABLI0335.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Disclosure; SEQ ID NO 25488; 21pp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 2309 AA;

Query Match 10.4%; Score 86.5; DB 22; Length 2309;

Best Local Similarity 27.4%; Pred. No. 8.5; Mismatches 55; Indels 17; Gaps 5;

Matches 34; Conservative 18; Mismatches 55; Indels 17; Gaps 5;

Qy 22 QNFSROEVGATGA---VVGAVQALFGKSGRVSMAIGAVLGLIGSKIGQSDQDK 78

Db 2124 QQQQQRQVGGNGSPSMALGRCGAGVSGSGN-----GGGGGAGAGSGVGGGNGV 2178

Qy 79 IKLNQSLKVKAGQVTRMRNPDGNSYSVEPVTRYORYNKQEROQYCRFPQOKAMIA 137

Db 2179 GSVGQS-----GGGGGQGYTPPIQRPNNYPOHP-----000000000000REDAAAAAVAA 2230





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:09:08 : Search time 1529.81 Seconds  
(without alignments)  
6775.784 Million cell updates/sec

Title: US-09-677-374-5  
Perfect score: 768  
Sequence: 1 atgcacgtgactctacaa.....aggtgattagcaccgcaaaa 768

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlinu:\*  
5: em\_estlov:\*  
6: em\_estlpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_huv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.4	5.7	551	9 AU037653	AU037653 AU037653
2	43	5.6	1100	12 CDS00FCD	AL070512 Drosophila
3	42.4	5.5	470	12 AZ799648	AZ799648 2M057B14
4	42.2	5.5	976	12 CDS04E5M	AL286627 Tetradon
5	42	5.5	238	10 BG240789	BG240789 OVI_38.CO
6	42	5.5	273	12 A0260501	A0260501 CITR1-El-
7	42	5.5	402	10 BG356895	BG356895 OV2_11.DO
8	42	5.5	419	10 BE353894	BE353894 DG1_11.El
9	42	5.5	428	10 BF176742	BF176742 EM1_4.F01
10	42	5.5	446	10 BG052290	BG052290 RH12_12_
11	42	5.5	500	10 BF586874	BF586874 FM1_31.E0
12	42	5.5	507	10 BG053591	BG053591 RH12_11_
13	42	5.5	531	10 BG053014	BG053014 RH12_16_
14	42	5.5	537	10 BG713844	BG713844 EM1_20.CO
15	42	5.5	561	10 BF587805	BF587805 FM1_40.GO
16	42	5.5	577	10 BG673858	BG673858 EM1_20.CO
17	42	5.5	582	10 BE356763	BE356763 DG1_12.B1

18	42	5.5	591	10 BE356800	BE356800 DG1_12.E1
19	42	5.5	597	10 BG102589	BG102589 RH12_34_
20	42	5.5	598	10 AM672446	AM672446 LG1_359_L0
21	42	5.5	600	10 BE360868	BE360868 DG1_67_A0
22	42	5.5	620	10 BF481524	BF481524 FM1_21_B0
23	42	5.5	649	10 BE355895	BE355895 DG1_11.El
24	41.8	5.4	238	9 AA352399	AA352399 EST60642
25	41.8	5.4	474	10 BM031893	BM031893 497456.MA
26	41.6	5.4	321	9 AM021656	AM021656 df26e07.y
27	41.6	5.4	361	9 AA653164	AA653164 ac52h10.s
28	41.6	5.4	459	10 B1493124	B1493124 df97h08.y
29	41.6	5.4	509	12 AZ447316	AZ447316 IM0244E04
30	41.6	5.4	556	10 BE395015	BE395015 601309414
31	41.6	5.4	597	10 B1492596	B1492596 df26e07.w
32	41.6	5.4	671	9 AM328753	AM328753 EST00024
33	41.6	5.4	742	10 BF792098	BF792098 602252539
34	41.6	5.4	878	10 BM449796	BM449796 AGENCOURT
35	41.6	5.4	958	12 CDS001FF	AL074914 Drosophila
36	41.6	5.4	1058	10 BM478067	BM478067 AGENCOURT
37	41.6	5.4	1079	10 BF792265	BF792265 602252768
38	41.6	5.4	1201	12 CDS0167M	AL106386 Drosophila
39	41.2	5.4	702	10 W27594	W27594 34h4 Human
40	41.2	5.4	737	10 BM166352	BM166352 EST568875
41	41	5.3	928	12 CDS00DKY	AL071865 Drosophila
42	40.6	5.3	942	12 CDS018GS	AL109318 Drosophila
43	40.4	5.3	548	10 BM328366	BM328366 PIC1_28.D
44	40.4	5.3	609	12 A2421257	A2421257 IM0199M21
45	40.4	5.3	856	12 A0740253	A0740253 HS_5505_A

#### ALIGNMENTS

RESULT 1  
LOCUS AU037653 551 bp mRNA linear EST 29-MAR-1999  
DEFINITION AU037653 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium  
ACCESSION AU037653  
VERSION AU037653.1 GI:3984406  
KEYWORDS EST.  
SOURCE Dictyostellium discoideum.  
ORGANISM Dictyostellium discoideum  
REFERENCE 1 (bases 1 to 551)  
Mori,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
Yoshino,R., Mitra,B.N., Pl.M., Sato,T., Takemoto,K., Yasukawa,H.,  
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.  
The Dictyostellium developmental cDNA project: generation and  
analysis of expressed sequence tags from the first-finger stage of  
development  
DNA Res. 5 (6), 335-340 (1998)  
MEDLINE 99156227  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp  
PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.  
FEATURES  
source location/Qualifiers  
1..551  
/organism="Dictyostellium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone\_lib="SSE138"  
/clone="Dictyostellium discoideum SS (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 266 a 63 c 39 g 183 t  
ORIGIN  
Query Match 5.7%; Score 43.4; DB 9; Length 551;  
Best Local Similarity 52.5%; Pred. No. 3;  
Matches 95; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

QY      32 CAGCACAACAACATCAATTACACATATCAAAATTTACTACACATCTGCAGTGATT 91
      34 CACACTATACAAATTAATTTAAATATCAATAAAAAATCGATTAATCAAAAAATTTAAATT 93
QY      92 TAAATTTAATGACGTAAAGTTAGATATTATTACACAGTGTGTACACAGACAA 151
      94 CAAAATCACATATTTTAAAAAAGAAATTTTAAATCAATAGTTACTTCATCAATCAAA 153
QY      152 CTTTCTGTGTGACATGCTGTCATTTATTAGGAATATGCTATGTTGTAACACTAGCA 211
      154 ATTCAATTAATTAATTGCAACTTTAATTTTACCAAAAAATGATGATTAATAATAA 213
QY      212 A 212
      214 A 214

RESULT 2
CNS00FCD 1100 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BAC3D18 of RPCI-98 library from Drosophila melanogaster (fruit
ACCESSION AL070512.1 GI:4950453
VERSION AL070512.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Genoscope.
          Direct Submission
          Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osoegawa and
          Aaron Mammoler in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw.sp, the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
          source
          1..1100
          /organism="Drosophila melanogaster"
          /db_xref="taxon:7227"
          /clone_lib="RPCI-98"
          /note="BAC3D18"
          /note="end : TET3"
BASE COUNT 356 a 100 c 111 g 137 t 396 others
ORIGIN
Query Match 5.6%; Score 43; DB 12; Length 1100;
Best local Similarity 33.7%; Pred. No. 5;
Matches 83; Conservative .6; Mismatches 157; Indels 0; Gaps 0;
QY      25 AACCAATTCAGCACAACAATCAATTTACACCAATATCAACATCTGCAGTGC 84
      104 AAAAAAAAAAAAAAAAAAAAAAAAAAANACATNNNNNATAATTNNACAATNTNANN 163
QY      85 ACGGATTTAATTTAATGACGTAAAGTTAGATATTATTACACAAGTGATGTACACAA 144
      104 AAAAAAAAAAAAAAAAAAAAAAAAAAANACATNNNNNATAATTNNACAATNTNANN 163

```

```

Db      164 CGASGGAANAANAAGGAANNANNNAAAAAHAAYAAANATTAAMAAAAA 223
QY      145 GGACAACCTTTCTGGTGTGACCATGCTGCTGCTATTATAGAAATAGCTATGTTGATAC 204
      224 CAGGACATTTTANNNNNANTNNNTNTNTNNNNANANNNANANNNANNNANNNAN 283
QY      205 ACTAGCAAGTGACAGCAACCTGTTAAAGAAACAGCAGCCCAACATCACTATGAT 264
      284 NAAAAAANNNANNNANNNNNNNNNNNNNNNNNANNCACACNNNNAAAAA 343
QY      265 ACATAT 270
      344 AAAKAT 349

RESULT 3
A2799648 470 bp DNA linear GSS 16-FEB-2001
LOCUS 2M0057B14F Mouse 10kb plasmid UGCG1M library Mus musculus genomic
DEFINITION clone UGCG2M0057B14 F, DNA sequence.
ACCESSION A2799648
VERSION A2799648
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 470)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beecorn,T., Duval,B., Hamil,C.,
          Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Kelly
          M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
          and Wright,D., Weiss,R.
          Mouse whole genome scaffolding with paired end reads from 10kb
          Plasmid Inserts
          Unpublished (2000)
          Contact: Robert B. Weiss
          University of Utah Genome Center
          Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
          84112, USA
          Tel: 801 385 5606
          Fax: 801 385 7177
          Email: ddunn@genetics.utah.edu
          Insert Length: 1000 Std Error: 0.00
          Plate: 0057 row: B column: 14
          Seq primer: CGTTGTAACGACGCGCAGT
          Class: plasmid ends
          High quality sequence stop: 470.
FEATURES
          source
          1..470
          /organism="Mus musculus"
          /strain="C57BL/6J"
          /db_xref="taxon:10090"
          /clone="UGCG2M0057B14"
          /clone_lib="Mouse 10kb plasmid UGCG1M library"
          /sex="Male"
          /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
          /note="Vector: PWD42ny. Purified genomic DNA from M.
          musculus C57BL/6J (male) was obtained from the Jackson
          Laboratory Mouse DNA Resource
          (http://www.jax.org/resources/documents/dnares/). The DNA
          was hydrodynamically sheared by repeated passage through a
          0.005 inch orifice at constant velocity. The sheared DNA
          was blunt end-repaired with T4 DNA polymerase and T4
          polynucleotide kinase. Adaptor oligonucleotides were
          ligated to the blunt ends in high molar excess. The
          adapted DNA was purified and size-selected for a 9.5 to
          10.5 kb range using preparative agarose gel
          electrophoresis. Vector DNA was prepared from a derivative
          of PWD42 (g14732114|9b|AP129072.1), a copy-number
          inducible derivative of plasmid R1. The vector was ligated
          with adaptors complementary to the insert adaptors and
          purified. The sheared, adapted mouse DNA was annealed to

```

BASE COUNT	ORIGIN
138 a	97 c 132 g 103 t

adaplored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match	5.5%	Score	42.4	DB	12	Length	470
Best Local Similarity	57.6%	Pred. No.	5				
Matches	76	Conservative	0	Mismatches	56	Indels	0
						Gaps	0

OY	2	TGTGCTGGAAATTTCACCACTCTAACCAATCAGCACAAACAACCTGAATTACCCCAATAA	61
Db	297	TGGTAGCTGTTTGTGTTCATCTACTGTTTACCACCAACCAACTTAACCCAAAACAAAA	356
OY	62	TCAAAATTACTAACACATCTGCAGTGATTTAAATTTAAATGACGTAAGAAGTTAGATATT	121
Db	357	TCAAAACTCCAATTAACCCAGCATTAACATAAATGACCATGAGCTCAATGTTGTAATTGG	416
OY	122	ATTACCAAGTG	133
Db	417	ATGTCATTGGTG	428

LOCUS	CNS04E5M	976 bp	DNA	Linear	GSS 21-MAY-2000
DEFINITION	Tetradon nigriviridis genome survey sequence T7 end of clone 103P02 of library G from Tetradon nigriviridis, genomic survey sequence.				

ACCESSION	AL286627	GI:8025084
VERSION	AL286627.1	
KEYWORDS	GSS: genome survey sequence.	
SOURCE	Tetraodon nigroviridis.	
ORGANISM	Tetraodon nigroviridis	

FEATURES	Location/Qualifiers
source	1. .976

BASE COUNT	243 a	151 c	137 g	327 t	118 others
ORIGIN					

Query Match	5.58;	Score 42.2;	DB 12;	Length 976;
Best Local Similarity	33.58;	Pred. NO. 7.5;		
Matches 116;	Conservative 61;	Mismatches 168;	Indels 1;	Gaps 1

Oy	11	AATTCACAGCTGTACAAATCGACCAACCAACCTCATTTACACCAATTAACAATTA	70
		:    :    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	888	WATTTATATATWTTTNTTAAANANAMAMATATAATAAAAATWMAWMWMTWTATATAAATTTTA	829
Oy	71	CTACACATCTGCACGTATTTAAATTTAAATGAGCTAAAGTTGATATATTATTCACAA	130
		:    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	828	TAAAAAAATTTWTTTNTTATTTTAAATWTTTAAATATAAAAA-TAMWMTTAMWTTATNTWT	770
Oy	131	GTGATGGTACACAAAGCAAACTTTCTGTGTGACCATCTGGTCATTTATTAGGAATA	190
		:    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	769	AWAAWAMATWTTWAAWMAAAATWTTAAWMTWMTATTAATATATWATATTATATWMAW	710
Oy	191	GCTATGTTGATPACCTAGCAAAAGTACAGCAAACTTGTTTAAAGAAACAGCAACCCAA	250
		:    :    :    :    :    :    :    :    :    :    :    :    :	
Db	709	TAAATTTTATTAATWMTNNATTTAAAAATAAATAAAMWMAWMAWMAWMAWMAWMAWMA	650
Oy	251	CATCAACCTATGATACATATCTGATCCGCTGCATATGCGTGCTGGCCGCGGACACT	310
		:    :    :    :    :    :    :    :    :    :    :    :    :    :	
Db	649	WMAWMAAAAKKKTTHHMMKKKKKKKKKKRGGGBKBVTTTIVBAGRGVBASGA	590
Oy	311	CTCTGATCATTTATCTGTGTTTCCGTGGGTGGCGCCAGCAACTT	356
		:    :    :    :    :    :    :    :    :    :    :    :    :	
Db	569	GGAGGACGGAGGAKAGCKYCTCTAGGGGGGATTTCCACACACTT	544

RESULT 5	LOCUS	DEFINITION
BG240789	238 bp	linear
BG240789	238 bp	EST 15-FEB-2001
OVI_38.C06.g1-A002	Ovary 1 (OVI)	Sorghum bicolor CDNA, mRNA sequence.

FEATURES	Location/Qualifiers
source	1. . 238

BASE COUNT	63 a	54 c	70 g	51 t
ORIGIN				

Query Match	5.5%;	Score 42;	DB 10;	Length 238;
Best Local Similarity	59.0%;	Pred. No. 4.8;		

Db	246	AA	247
RESULT 7			
LOCUS	BG356895		
DEFINITION	OV2.11.D01.g1_A002 Ovary 2 (OV2)	402 bp	mRNA linear EST 06-MAR-2001
ACCESSION	BG356895		
VERSION	BG356895.1		
KEYWORDS	EST.		
SOURCE	Sorghum bicolor		
ORGANISM	Sorghum bicolor		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.		
AUTHORS	1 (bases 1 to 402)		
TITLE	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.		
JOURNAL	An EST database from Sorghum: ovaries of varying immature stages		
COMMENT	Unpublished (2000) Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmprratt@uga.edu Seq primer: PolyTMix High quality sequence start: 47 High quality sequence stop: 391 POLYA=No		
FEATURES			
Source	Location/Qualifiers		
	1..402		
	/organism="Sorghum bicolor"		
	/db_xref="taxon:4558"		
	/clone_lib="Ovary 2 (OV2)"		
	/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from polyA RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."		
BASE COUNT	107 a 82 c 117 g 96 t		
ORIGIN			
Query Match	5.5%; Score 42; DB 10; Length 402;		
Best Local Similarity	59.0%; Pred. No. 5.9;		
Matches	72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;		
OY	345 GCCCAGAACTTCAGCCGACGAGAACTTGGCGGCGCCACCGGTCGTTGGCGCGTGT 404		
Db	140 CGGGAGAGAGCTGCACGACGAGGAGTCGACGAGATGATCCGAGAGCGTACGTCATGG 199		
OY	405 TGGCGGCGAGCGTTCGTTAAGGCGTCGTGCTGTGTCATGCGCATCGCGGTCGCGT 464		
Db	200 CGAGGCGCCATCACTATGAGAAGATTGTTAAGATTATGATGCGCAAGTGAAGAGCGGT 259		
OY	465 TC 466		
Db	260 CC 261		
RESULT 8			
LOCUS	BE355894		
DEFINITION	DG1.11.E11.g1_A002 Dark Grown 1 (DG1)	419 bp	mRNA linear EST 20-JUL-2000
ACCESSION	BE355894		
VERSION	BE355894.1		
KEYWORDS	EST.		
SOURCE	Sorghum bicolor		
ORGANISM	Sorghum bicolor		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		





**RESULT 13**  
**LOCUS** BG053014 531 bp mRNA linear EST 25-JAN-2001  
**DEFINITION** RH12\_16.D04.g1\_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA sequence.  
**ACCESSION** BG053014  
**VERSION** BG053014.1 GI:12508270  
**KEYWORDS** EST.  
**SOURCE** Sorghum propinquum.  
**ORGANISM** Sorghum propinquum  
**REFERENCE** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
**AUTHORS** 1 (bases 1 to 531)  
 Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt,L.H.  
**TITLE** An EST database from Sorghum: Sorghum propinquum rhizomes  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu  
 Seq primer: PolyTmix  
 High quality sequence start: 42  
 High quality sequence stop: 527  
 POLYA-No.

**FEATURES**  
 source location/Qualifiers  
 1..531  
 /organism="Sorghum propinquum"  
 /db\_xref="taxon:132711"  
 /clone\_lib="Rhizome2 (RH122)"  
 /note="Organ: Rhizomes; Vector: Bluescript II from lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."  
**BASE COUNT** 122 a 101 c 152 g 154 t 2 others

**Query Match** 5.5%; Score 42; DB 10; Length 531;  
**Best Local Similarity** 59.0%; Pred. No. 6.6;  
**Matches** 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

**QY** 345 CGCCAGACACTTCAGCCGCGAGAAATTGGCGCGCCACCGCGTGGTGGCGGGT 404  
 11 11111 11 11111 11 11 11 11111 11 11 11 11 11 11 11 11  
**DB** 121 CGGCGAGACAGTGCACGACGAGAGTGCGACGAGATATCCGTAGGCTGACGTGATCG 180  
 11 11111 11 11111 11 11 11 11111 11 11 11 11 11 11 11 11  
**QY** 405 TGCCGGCCAGCTGTTGGTAAAGGCTCTGTCGTGTCGATGCGCATCGCGGTCGGGT 464  
 1111111 11111 11111 11 11 11 111111 1111111 1111111  
**DB** 181 CGACGGCCAGATCAACATATGAAGATTGTTAAGATTATATATGCGCAAGTAGAGAGCGGT 240  
 1111111 11111 11111 11 11 11 111111 1111111 1111111  
**QY** 465 TC 466  
 1  
**DB** 241 CC 242

**RESULT 14**  
**LOCUS** BG713844 537 bp mRNA linear EST 08-MAY-2001  
**DEFINITION** EM1\_20\_G08.g2\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.  
**ACCESSION** BG713844  
**VERSION** BG713844.1 GI:14007794  
**KEYWORDS** EST.  
**SOURCE** Sorghum.  
**ORGANISM** Sorghum bicolor  
**REFERENCE** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

**REFERENCE** 1 (bases 1 to 537)  
**AUTHORS** Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.  
**TITLE** An EST database from Sorghum: developing embryos  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu  
 Sequences have been trimmed to exclude Polya, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.  
 Seq primer: PolyTmix  
 High quality sequence start: 62  
 High quality sequence stop: 536  
 POLYA-No.

**FEATURES**  
 source location/Qualifiers  
 1..537  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Embryo 1 (EM1)"  
 /note="Organ: Embryos germinated for 24 hr; Vector: Bluescript II from lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."  
**BASE COUNT** 118 a 98 c 152 g 169 t

**Query Match** 5.5%; Score 42; DB 10; Length 537;  
**Best Local Similarity** 59.0%; Pred. No. 6.6;  
**Matches** 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

**QY** 345 CGCCAGACACTTCAGCCGCGAGAAATTGGCGCGCCACCGCGTGGTGGCGGGT 404  
 11 11111 11 11111 11 11 11 11111 11 11 11 11 11 11 11 11  
**DB** 92 CGGCGAGACAGTGCACGACGAGAGTGCGACGAGATATCCGTAGGCTGACGTGATCG 151  
 11 11111 11 11111 11 11 11 11111 11 11 11 11 11 11 11 11  
**QY** 405 TGCCGGCCAGCTGTTGGTAAAGGCTCTGTCGTGTCGATGCGCATCGCGGTCGGGT 464  
 1111111 11111 11111 11 11 11 111111 1111111 1111111  
**DB** 152 CGACGGCCAGATCAACATATGAAGATTGTTAAGATTATATGCGCAAGTAGAGAGCGGT 211  
 1111111 11111 11111 11 11 11 111111 1111111 1111111  
**QY** 465 TC 466  
 1  
**DB** 212 CC 213

**RESULT 15**  
**LOCUS** BF587805 561 bp mRNA linear EST 12-DEC-2000  
**DEFINITION** EM1\_40\_G05.g1\_A003 Floral-Induced Meristem 1 (FMI) Sorghum propinquum cDNA, mRNA sequence.  
**ACCESSION** BF587805  
**VERSION** BF587805.1 GI:11680115  
**KEYWORDS** EST.  
**SOURCE** Sorghum propinquum.  
**ORGANISM** Sorghum propinquum  
**REFERENCE** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.  
**AUTHORS** 1 (bases 1 to 561)  
 Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt,L.H.  
**TITLE** An EST database from Sorghum: floral-induced meristems  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Cordonnier-Pratt MM  
 Department of Botany  
 The University of Georgia  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 542 1805  
 Email: mmp@prattuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTmix  
High quality sequence

High quality sequence start: 81

High quality sequence stop: 561  
 POTVA=No

POLYA=NO.

## FEATURES

### Location/Qualifiers

### Source

1. .561

```
/organism="Sorghum propinquum"  
/db_xref="taxon:132711"
```

```
/db_xref="taxon:132711"
```

/db\_xref="taxon:132711"

```
/clone_lib="Floral-Induced Meristem 1 (FM1)"
```

pbjescrpt II from Lambda Zap II: Site 1: ybcr

PBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2:

ECORI; mature plants were placed in a growth chamber for

15 days with 16 hr darkness and 8 hr light (43°C).

15 days with 16 hr darkness and 8 hr light (flowering i

induced by short-day conditions): 16 days after being

Induced by short-day conditions); 16 days after being

returned to the greenhouse under natural long days during

late April/early May. meristems were harvested during

late April/early May, meristems were harvested. The

ORIGIN

### Query Match

Score 42; DB 10; Length 561;

Best Local Similarity 59.0%; Pred. No. 6,8;

Matches	72;	Conservative	0;	Mismatches	50;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

345 CGCCGAACTTCAGCCGCCAGGAAGTTGGCGCGGCCACCGGTGCGGTTGTGGCGGTGT 404

Db 202 CGCGGAGAGCTGACCGACGAGGAGGTCGACGAGATGATCCGTGAGGCTGACGTCGATGC 261

405 TGGCGGCCAGCTGTTCCGGTAAGGCTCTGGTCCTGTCTCGATGGCCATCGGGGTGGGT 464

Db 262 CGACGGCCAGATCACTATGAGAGTTTGTAAAGTTATGATGCCCAAGTGAGGAGCCGT 321

QY	465	TC	466
----	-----	----	-----

Db 322 CC 323

Search completed: October 27, 2002, 18:34:59  
Job time : 1533.81 secs

Job time : 1533.81 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 16:06:08 : Search time 56.152 Seconds  
(without alignments)  
3359.570 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atgcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/1na/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/1na/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/1na/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/1na/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270.2	35.2	486	1	US-08-048-164A-1 Sequence 1, Appl
2	270.2	35.2	486	1	US-08-048-164A-3 Sequence 3, Appl
3	270.2	35.2	486	1	US-08-460-462-1 Sequence 1, Appl
4	270.2	35.2	486	1	US-08-460-462-3 Sequence 3, Appl
5	270.2	35.2	486	1	US-08-460-457-1 Sequence 1, Appl
6	270.2	35.2	486	1	US-08-460-457-3 Sequence 3, Appl
7	270.2	35.2	486	1	US-08-460-458-1 Sequence 1, Appl
8	270.2	35.2	486	1	US-08-460-458-3 Sequence 3, Appl
9	270.2	35.2	486	2	US-08-460-455-1 Sequence 1, Appl
10	270.2	35.2	486	2	US-08-460-455-3 Sequence 3, Appl
11	270.2	35.2	486	2	US-08-330-394A-1 Sequence 1, Appl
12	270.2	35.2	486	2	US-08-330-394A-3 Sequence 3, Appl
13	270.2	35.2	486	3	US-09-006-636-5 Sequence 5, Appl
14	270.2	35.2	486	4	US-09-006-632-5 Sequence 21, Appl
15	270.2	35.2	486	4	US-09-277-716-21 Sequence 9, Appl
16	70.2	9.1	1482	4	US-09-198-956-9 Sequence 11, Appl
17	70.2	9.1	1482	4	US-09-198-956-11 Sequence 31, Appl
18	40	5.2	2645	4	US-08-471-044-6 Sequence 6, Appl
19	40	5.2	2645	4	US-08-471-044-6 Sequence 6, Appl
20	40	5.2	2645	4	US-09-073-898-31 Sequence 1, Appl
21	39.4	5.1	2004	2	US-08-471-044-6 Sequence 6, Appl
22	39.4	5.1	2004	2	US-08-471-044-6 Sequence 6, Appl
23	39.4	5.1	2004	2	US-08-471-044-6 Sequence 6, Appl
24	39.4	5.1	2004	2	US-08-471-044-6 Sequence 6, Appl
25	39.4	5.1	2004	2	US-08-471-044-6 Sequence 6, Appl
26	39.4	5.1	2004	2	US-08-471-044-6 Sequence 6, Appl
27	39.4	5.1	2004	3	US-09-300-529-6 Sequence 6, Appl

28	39.4	5.1	2655	1	US-08-471-033-4 Sequence 4, Appl
29	39.4	5.1	2655	2	US-08-471-044-4 Sequence 4, Appl
30	39.4	5.1	2655	2	US-08-463-483A-4 Sequence 4, Appl
31	39.4	5.1	2655	2	US-08-471-046A-4 Sequence 4, Appl
32	39.4	5.1	2655	2	US-08-470-566B-4 Sequence 4, Appl
33	39.4	5.1	2655	2	US-08-469-334-4 Sequence 4, Appl
34	39.4	5.1	2655	3	US-09-300-529-4 Sequence 4, Appl
35	39.4	5.1	4041	1	US-08-471-033-22 Sequence 22, Appl
36	39.4	5.1	4041	2	US-08-471-044-22 Sequence 22, Appl
37	39.4	5.1	4041	2	US-08-463-483A-22 Sequence 22, Appl
38	39.4	5.1	4041	2	US-08-471-046A-22 Sequence 22, Appl
39	39.4	5.1	4041	2	US-08-470-566B-22 Sequence 22, Appl
40	39.4	5.1	4041	2	US-08-469-334-22 Sequence 22, Appl
41	39.4	5.1	4041	3	US-09-300-529-22 Sequence 22, Appl
42	39.4	5.1	6049	1	US-08-471-033-1 Sequence 1, Appl
43	39.4	5.1	6049	2	US-08-471-044-1 Sequence 1, Appl
44	39.4	5.1	6049	2	US-08-463-483A-1 Sequence 1, Appl
45	39.4	5.1	6049	2	US-08-471-046A-1 Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-048-164A-1  
Sequence 1, Application US/08048164A  
Patent No. 5496934  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
City: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/048,164A  
FILING DATE: 14-Apr-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-048-164A-1  
Query Match 35.2%; Score 270.2; DB 1; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1

Matches 272: Conserve

Query Match	35.2%	Score 270.2;	DB 1;	Length 486;
Best Local Similarity	98.9%;	Pred. No. 3,8e-73;		
Matches 272;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1	ATGTGATTCGATTCTTACACACTCTTACAACTAATGAGACAAACAACTCAATTTACACCAATA	60	

Db 16 ATGCACTGAATTTTACAACTCTAACAAATGACCAACAAACCTCAATTACACCAATA 75  
Qy 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTATAT 120  
Db 76 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTATAT 135  
Qy 121 TATTACAAAGTGTGTGTACACAAAGCAAACTTCTGTGTGACCAATGCTGGTGCATTA 180  
Db 136 TATTACAAAGTGTGTGTACACAAAGCAAACTTCTGTGTGACCAATGCTGGTGCATTA 195  
Qy 181 TTAGAAATAGCTATGTTGTAACACTAGCAAGTACAGCAAACTTCTGTTAAAGAACA 240  
Db 196 TTAGAAATAGCTATGTTGTAACACTAGCAAGTACAGCAAACTTCTGTTAAAGAACA 255  
Qy 241 GCAAGCCCAACATCACCCTATGATACATATCTGCA 275  
Db 256 GCAAGCCCAACATCACCCTATGATACATATCTGCA 290

## RESULT 4

US-08-460-462-3/C  
Sequence 3, Application US/08460462  
Patent No. 5670623  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shplegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,462  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-460-462-3

Query Match 35.2%; Score 270.2; DB 1: Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 ATGCACTGAATTTTACAACTCTAACAAATGACCAACAAACCTCAATTACACCAATA 60  
|||||

Db 471 ATGCACTGAATTTTACAACTCTAACAAATGACCAACAAACCTCAATTACACCAATA 412  
Qy 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTATAT 120  
Db 411 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAGTATAT 352  
Qy 121 TATTACAAAGTGTGTGTACACAAAGCAAACTTCTGTGTGACCAATGCTGGTGCATTA 180  
Db 351 TATTACAAAGTGTGTGTACACAAAGCAAACTTCTGTGTGACCAATGCTGGTGCATTA 292  
Qy 181 TTAGAAATAGCTATGTTGTAACACTAGCAAGTACAGCAAACTTCTGTTAAAGAACA 240  
Db 291 TTAGAAATAGCTATGTTGTAACACTAGCAAGTACAGCAAACTTCTGTTAAAGAACA 232  
Qy 241 GCAAGCCCAACATCACCCTATGATACATATCTGCA 275  
Db 231 GCAAGCCCAACATCACCCTATGATACATATCTGCA 197

## RESULT 5

US-08-460-457-1  
Sequence 1, Application US/08460457  
Patent No. 5719044  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shplegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,457  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-457-1

Query Match 35.2%; Score 270.2; DB 1: Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTCCTACAACTCTACAAATCAGCAACAACTCAATTCACCAATA 60  
Db 16 ATGTCAGTTGAATTCCTACAACTCTACAAATCAGCAACAACTCAATTCACCAATA 75  
QY 61 ATCAAAATTTACTACACACTCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 120  
Db 76 ATCAAAATTTACTACACACTCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 135  
QY 121 TATTACACAGTGATGTCACACAGCAAGCAAACTTCTGGTGTCACATGCTGTGCATTA 180  
Db 136 TATTACACAGTGATGTCACACAGCAAGCAAACTTCTGGTGTCACATGCTGTGCATTA 195  
QY 181 TTAGAAATTAAGCTATGTTGATTAACACTAGCAAGTAGACACAACTCGTTAAGAACA 240  
Db 196 TTAGAAATTAAGCTATGTTGATTAACACTAGCAAGTAGACACAACTCGTTAAGAACA 255  
QY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275  
Db 256 GCAAGCCCAACATCAACCTATGATATATGTTGA 290

RESULT 6  
US-08-460-457-3/C  
Sequence 3, Application US/08460457  
Patent No. 5719044  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpilel, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,457  
FILING DATE: Concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-460-457-3

Query Match 35.2%; Score 270.2; DB 1; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGTTGAATTCCTACAACTCTACAAATCAGCAACAACTCAATTCACCAATA 60  
Db 471 ATGTCAGTTGAATTCCTACAACTCTACAAATCAGCAACAACTCAATTCACCAATA 412  
QY 61 ATCAAAATTTACTACACACTCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 120  
Db 411 ATCAAAATTTACTACACACTCTGACAGTGATTTAAATTTAAATGACGTAAGTAGATAT 352  
QY 121 TATTACACAGTGATGTCACACAGCAAGCAAACTTCTGGTGTCACATGCTGTGCATTA 180  
Db 351 TATTACACAGTGATGTCACACAGCAAGCAAACTTCTGGTGTCACATGCTGTGCATTA 292  
QY 181 TTAGAAATTAAGCTATGTTGATTAACACTAGCAAGTAGACACAACTCGTTAAGAACA 240  
Db 291 TTAGAAATTAAGCTATGTTGATTAACACTAGCAAGTAGACACAACTCGTTAAGAACA 232  
QY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275  
Db 231 GCAAGCCCAACATCAACCTATGATATATGTTGA 197

RESULT 7  
US-08-460-458-1  
Sequence 1, Application US/08460458  
Patent No. 5738984  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,458  
FILING DATE: Concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-458-1

Query Match 35.2%; Score 270.2; DB 1; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 16 ATGTCAAGTGAATTTTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 75  
QY 61 ATCAAAATTTACTACACATCTGACAGTGAATTTAATTAATGACGTAAGAGTAT 120  
Db 76 ATCAAAATTTACTACACATCTGACAGTGAATTTAATTAATGACGTAAGAGTAT 135  
QY 121 TATTAACAAGTGTGTGTACACAAGCAAACTTGTGTGTGACCAATGCTGTGATTA 180  
Db 136 TATTAACAAGTGTGTGTACACAAGCAAACTTGTGTGTGACCAATGCTGTGATTA 195  
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTGTTAAAGAAACA 240  
Db 196 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTGTTAAAGAAACA 255  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGA 275  
Db 256 GCAAGCCCAACATCAACCTATGATACATATCTGA 290

RESULT 8  
US-08-460-458-3/C  
: Sequence 3, Application US/08460458  
: Patent No. 5738984  
: GENERAL INFORMATION:  
: APPLICANT: Shoseyov, Oded  
: TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN  
: NUMBER OF SEQUENCES: 21  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: PENNIE & EDMONDS  
: STREET: 1155 Avenue of the Americas  
: CITY: New York  
: STATE: New York  
: COUNTRY: U.S.A.  
: ZIP: 10036  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/460,458  
: FILING DATE: concurrently herewith  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/048,164  
: FILING DATE: 14-APR-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Mistrock, S. Leslie  
: REGISTRATION NUMBER: 18,872  
: REFERENCE/DOCKET NUMBER: 7809-007  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (212) 790-9090  
: TELEFAX: (212) 869-8864/9741  
: TELEX: 66141 PENNIE  
: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 486 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: unknown  
: MOLECULE TYPE: DNA  
: US-08-460-458-3

Query Match 35.2%; Score 270.2; DB 1; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAAGTGAATTTTACTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 60  
Db 471 ATGTCAAGTGAATTTTACTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 412  
QY 61 ATCAAAATTTACTACACATCTGACAGTGAATTTAATTAATGACGTAAGAGTAT 120  
|||||

Db 411 ATCAAAATTTACTACACATCTGACAGTGAATTTAATTAATGACGTAAGAGTAT 352  
QY 121 TATTAACAAGTGTGTGTACACAAGCAAACTTGTGTGTGACCAATGCTGTGATTA 180  
Db 351 TATTAACAAGTGTGTGTACACAAGCAAACTTGTGTGTGACCAATGCTGTGATTA 292  
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTGTTAAAGAAACA 240  
Db 291 TTAGGAATAGCTATGTTGATACACTAGCAAGCAAACTTGTGTTAAAGAAACA 232  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGA 275  
Db 231 GCAAGCCCAACATCAACCTATGATACATATCTGA 197

RESULT 9  
US-08-460-455-1  
: Sequence 1, Application US/08460455  
: Patent No. 5837814  
: GENERAL INFORMATION:  
: APPLICANT: Shoseyov, Oded  
: APPLICANT: Shpiegl, Itai  
: APPLICANT: Goldstein, Marc A.  
: APPLICANT: Dol, Roy H.  
: TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS  
: NUMBER OF SEQUENCES: 21  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: PENNIE & EDMONDS  
: STREET: 1155 Avenue of the Americas  
: CITY: New York  
: STATE: New York  
: COUNTRY: U.S.A.  
: ZIP: 10036  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/460,455  
: FILING DATE: concurrently herewith  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/048,164  
: FILING DATE: 14-APR-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Mistrock, S. Leslie  
: REGISTRATION NUMBER: 18,872  
: REFERENCE/DOCKET NUMBER: 7809-009  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (212) 790-9090  
: TELEFAX: (212) 869-8864/9741  
: TELEX: 66141 PENNIE  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 486 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: unknown  
: MOLECULE TYPE: DNA  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 1..486  
: US-08-460-455-1

Query Match 35.2%; Score 270.2; DB 2; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.8e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAAGTGAATTTTACTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 60  
Db 16 ATGTCAAGTGAATTTTACTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 75  
|||||

16 ATGTCAGTTGAATTTTACACACTCTACAAT

QY	61	ATCAAAATTTAGTAAACACACTTGTGACAGCGATTTAAATTTAAATGACGTAAGTAAAGTTAGTAT	120
QY	81	ATCAAAATTTAGTAAACACACTTGTGACAGCGATTTAAATTTAAATGACGTAAGTAAAGTTAGTAT <td>140</td>	140
QY	121	TATTACACAGTAGTGTACACAAGACAACCTTTCTGTGTGTGACCACTGCTGTGATTA	180

Db 141 TATTACAAAGTGTATGATACAAAGCAAACTTCTGTGTGACCATGCTGTCATTA 200  
OY 181 TTAGGAATAGCTATGTATGATACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACA 240  
Db 201 TTAGGAATAGCTATGTATGATACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACA 260  
OY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275  
Db 261 GCAAGCCCAACATCAACCTATGATACATATCTGGA 295

## RESULT 14

US-09-006-632-5  
; Sequence 5, Application US/09006632  
; Patent No. 6184440  
; GENERAL INFORMATION:  
; APPLICANT: Shoseyov, Oded  
; APPLICANT: Shani, Ziv  
; APPLICANT: Shpigel, Eral  
; TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED  
; TITLE OF INVENTION: MORPHOLOGY  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/006,632  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7809-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 499 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-09-006-632-5

Query Match 35.2%; Score 270.2; DB 4; Length 499;  
Best Local Similarity 98.9%; Pred. No. 3.9e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGATCTCACTGCTAACAATCGACAAACAAACAACTTAATACACCATTA 60  
Db 21 ATGTCAGTTGATCTCACTGCTAACAATCGACAAACAAACAACTTAATACACCATTA 80  
OY 61 ATCAAAATTTACTTAACATCTGACAGTATTTAAATTAATGACGTAAGTTAGATAT 120  
Db 81 ATCAAAATTTACTTAACATCTGACAGTATTTAAATTAATGACGTAAGTTAGATAT 140  
OY 121 TATTACCAAGTATGATGATACAAAGCAAACTTCTGTGTGACCATGCTGTGATTA 180  
Db 141 TATTACCAAGTATGATGATACAAAGCAAACTTCTGTGTGACCATGCTGTGATTA 200  
OY 181 TTAGGAATAGCTATGTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAGAAACA 240

Db 201 TTAGGAATAGCTATGTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAGAAACA 260  
OY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275  
Db 261 GCAAGCCCAACATCAACCTATGATACATATCTGGA 295

## RESULT 15

US-09-277-716-21  
; Sequence 21, Application US/09277716A  
; Patent No. 6232107  
; GENERAL INFORMATION:  
; APPLICANT: Bryan, Bruce  
; APPLICANT: Szent-Gyorgyi, Christopher  
; APPLICANT: PROLUME, LTD.  
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE  
; CURRENT APPLICATION NUMBER: US/09-03-26  
; EARLIER FILING DATE: 1999-03-26  
; EARLIER APPLICATION NUMBER: 60/102,939  
; EARLIER FILING DATE: 1998-10-01  
; EARLIER APPLICATION NUMBER: 60/089,367  
; EARLIER FILING DATE: 1998-06-15  
; EARLIER APPLICATION NUMBER: 60/079,624  
; EARLIER FILING DATE: 1998-03-27  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 1146  
; TYPE: DNA  
; ORGANISM: Artificial Sequence: fusion construct  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1146)  
; FEATURE:  
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia lucifera  
US-09-277-716-21

Query Match 35.2%; Score 270.2; DB 4; Length 1146;  
Best Local Similarity 98.9%; Pred. No. 6.2e-73;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGATCTCACTGCTAACAATCGACAAACAAACAACTTAATACACCATTA 60  
Db 1 ATGTCAGTTGATCTCACTGCTAACAATCGACAAACAAACAACTTAATACACCATTA 60  
OY 61 ATCAAAATTTACTTAACATCTGACAGTATTTAAATTAATGACGTAAGTTAGATAT 120  
Db 61 ATCAAAATTTACTTAACATCTGACAGTATTTAAATTAATGACGTAAGTTAGATAT 120  
OY 121 TATTACCAAGTATGATGATACAAAGCAAACTTCTGTGTGACCATGCTGTGATTA 180  
Db 121 TATTACCAAGTATGATGATACAAAGCAAACTTCTGTGTGACCATGCTGTGATTA 180  
OY 181 TTAGGAATAGCTATGTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAGAAACA 240  
Db 181 TTAGGAATAGCTATGTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAGAAACA 240  
OY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275  
Db 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275

Search completed: October 27, 2002, 20:15:44  
Job time : 61.152 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 14:35:03 : Search time 194.984 Seconds

(without alignments)  
6762.541 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atgcacgttgatctctacaa.....agtgatcagcagcaaaaa 768

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

```
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	100.0	768	22	AAF86248
2	483	62.9	483	22	AAF86247
3	295	38.4	486	22	AAF86246
4	295	38.4	489	22	AAH79040
5	271.8	35.4	573	22	AAH11043
6	271.8	35.4	1030	22	AAH11044
7	270.2	35.2	486	15	AAQ72917
8	270.2	35.2	486	20	AAV74072
9	270.2	35.2	499	20	AAH24930

10	270.2	35.2	507	22	AAH11042	Clostridium cellui
11	270.2	35.2	984	22	AAH11046	Chimeric S peptide
12	270.2	35.2	1146	20	AAH227550	Gaussia luciferase
13	270.2	35.2	1146	24	AAH222201	Gaussia species CB
14	270.2	35.2	1288	22	AAH11045	Clostridium cellui
15	118	15.4	118	22	AAF86254	PCR primer #4 used
16	108.4	14.1	110	22	AAF86252	PCR primer #2 used
17	102	13.3	102	22	AAF86255	PCR primer #5 used
18	92.4	12.0	94	22	AAF86253	PCR primer #3 used
19	87	11.3	111	22	AAF86251	PCR primer #1 used
20	72.4	9.4	110	22	AAF86256	PCR primer #6 used
21	70.2	9.1	1438	21	AAZ45336	DNA encoding a man
22	70.2	9.1	1482	20	AAH90978	DNA encoding pecta
23	70.2	9.1	1482	20	AAH231562	Pectate lyase CPD
24	70.2	9.1	5562	19	AAF86625	C. thermocellum Ci
25	45.8	6.0	8246	24	ABH32202	Human immune syste
26	44.4	5.8	10595	22	AAH46541	Tumour suppressor
27	44	5.7	3632	24	ABA94342	Canine cyclooxygen
28	42	5.5	866	22	AAH94068	Human neuroblastom
29	41.8	5.4	12409	24	AAH63312	Chemically pretrea
30	41.6	5.4	390	22	AAF66075	Novel human polyu
31	41.6	5.4	1662	22	AAH59301	Human polynucleoti
32	41.6	5.4	1900	22	AAH93857	Human stomach canc
33	41.6	5.4	1900	22	AAH15558	Human stomach canc
34	41.6	5.4	2364	21	AAH23323	Human cDNA sequenc
35	41.6	5.4	2447	22	AAH161087	Human secreted pro
36	41.6	5.4	3389	22	AAH93892	Human polynucleoti
37	41.6	5.4	3389	22	AAH16279	Human stomach canc
38	41.2	5.4	270	21	AAH07093	Human immune syste
39	41.2	5.4	12601	24	ABH34206	Human immune syste
40	41.2	5.4	17934	20	AAH33718	Human gene express
41	41	5.3	759	24	AAH216548	Human polynucleoti
42	40.8	5.3	434	22	AAH190053	Human secreted pro
43	40.8	5.3	1104	22	AAH33253	Human immune syste
44	40.8	5.3	9707	24	ABH33420	Human hypocrerin r
45	40.8	5.3	168575	22	AAH21613	

#### ALIGNMENTS

```
RESULT 1
ID AAF86248 standard; DNA: 768 BP.
XX AAF86248;
AC AAF86248;
XX 11-JUL-2001 (first entry)
DT 11-JUL-2001 (first entry)
XX 11-JUL-2001 (first entry)
DE DNA sequence of c17e2 ospa construct with N-terminal fusion partner.
XX Polkiothermic fish; piscirickettsia salmonis; rickettsial pathogen;
KW vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;
KM SRS; 17E2; fusion construct; ds.
XX SRS; 17E2; fusion construct; ds.
OS Piscirickettsia salmonis.
OS Synthetic.
XX Synthetic.
XX Key Location/Qualifiers
FH CDS 1..768
FT /*tag= a
FT /*tag= b
FT /*tag= "DNA encoding undefined N-terminal fusion partner"
FT /*tag= "c17e2 ospa with N-terminal fusion partner"
FT /*tag= "No stop codon is given"
FT /*tag= "Optimised Ospa construct c17e2"
XX CA2281913-A1.
```

PD 17-MAR-2001.  
 XX 17-SEP-1999; 99CA-2281913.  
 PF 17-SEP-1999; 99CA-2281913.  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 XX (KAYW/) KAY W. W.  
 PA (BURI/) BURIAN J.  
 PA (KUZV/) KUZV M. A.  
 XX  
 PI Kay WW, Burian J, Kuzk MA;  
 XX  
 DR MPI: 2001-316844/34.  
 DR P-PSDB: AAB81127.  
 XX  
 PT Method for protecting poikilothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis -  
 XX  
 PS Example 4; Fig 5; 35pp; English.  
 XX  
 .CC This invention relates to a method for the protection against infection  
 .CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of a  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly poikilothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents P. salmonis Ospa DNA termed C17E2 optimised for  
 CC expression in Escherichia coli fused to DNA encoding an undefined  
 CC N-terminal fusion partner. The protein encoded by this fusion construct  
 CC is used in a vaccine to create an anti-Ospa antibody response.  
 XX  
 SQ Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other;  
 Query Match 100.0%; Score 768; DB 22; Length 768;  
 Best Local Similarity 100.0%; Pred. No. 1e-214;  
 Matches 768; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGTGAGTTGAAATCTTCACTAATCACTTACAAATCAGCACAACAACTCAATTACACCAATA 60  
 DB 1 ATGTGAGTTGAAATCTTCACTAATCACTTACAAATCAGCACAACAACTCAATTACACCAATA 60  
 QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120  
 DB 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTTAAATGACGTAAAGTTAGATAT 120  
 QY 121 TATTACACAAGTGTGATGATGACAGCAAGCAAACTTCTGGGTGATGACCATGCTGGTGCATTA 180  
 DB 121 TATTACACAAGTGTGATGATGACAGCAAGCAAACTTCTGGGTGATGACCATGCTGGTGCATTA 180  
 QY 181 TTAGGAATATAGCTATGTTGATTAACACTAGCAAAAGTGACACAACTCGTTAAAGAAACA 240  
 DB 181 TTAGGAATATAGCTATGTTGATTAACACTAGCAAAAGTGACACAACTCGTTAAAGAAACA 240  
 QY 241 GCAAGCCCAATCAACCTATGATATCTGATCTGATCCGTCATATGCTGTTGCTTG 300  
 DB 241 GCAAGCCCAATCAACCTATGATATCTGATCTGATCCGTCATATGCTGTTGCTTG 300  
 QY 301 CAGGGCAGCTCTCTGATCATTTATCTGTTTCTGTTGGGTGGCCGCGCAAGACTCAGC 360  
 DB 301 CAGGGCAGCTCTCTGATCATTTATCTGTTTCTGTTGGGTGGCCGCGCAAGACTCAGC 360  
 QY 361 CGCCAGAGATTTGGCGCGCCGCGGTGTTGGCGGGTTTCCGCCAGACTGTTTC 420  
 DB 361 CGCCAGAGATTTGGCGCGCCGCGGTGTTGGCGGGTTTCCGCCAGACTGTTTC 420  
 QY 421 GGTAAAGGCTCTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 DB 421 GGTAAAGGCTCTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 QY 481 GGTCTTAAATCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540

DB 481 GGTCTTAAATCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 QY 541 GAAAGAGTAAAGCCGCGCAGGTTACTGTTGGCGTAATTCGGACACCGGTAACAGCTAC 600  
 DB 541 GAAAGAGTAAAGCCGCGCAGGTTACTGTTGGCGTAATTCGGACACCGGTAACAGCTAC 600  
 QY 601 TCTGTGGACCGGTTGCGACCTACACGCTTACAAACAAGAGAAACCGCTCAGCAGTAC 660  
 DB 601 TCTGTGGACCGGTTGCGACCTACACGCTTACAAACAAGAGAAACCGCTCAGCAGTAC 660  
 QY 661 TGCCGCAATTTTCAGCAAAAGCCATGATGCGCAGCTGAGAAACGAAATCTACAGCACC 720  
 DB 661 TGCCGCAATTTTCAGCAAAAGCCATGATGCGCAGCTGAGAAACGAAATCTACAGCACC 720  
 QY 721 GCGTGCCCTCAGCCGCGATGCGCGCTGCGCAGGTGATAGCAACCGAAAAA 768  
 DB 721 GCGTGCCCTCAGCCGCGATGCGCGCTGCGCAGGTGATAGCAACCGAAAAA 768

RESULT 2  
 AAF86247  
 ID AAF86247 standard; DNA; 483 BP.  
 XX  
 AC AAF86247;  
 XX  
 DT 11-JUL-2001 (first entry)  
 XX  
 DE DNA sequence of E. coli optimised ospa gene 17E2.  
 XX  
 KW Poikilothermic fish: Piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; OSPA; salmonid rickettsial septicemia; rickettsial disease;  
 KM SRS; 17E2; ds.  
 XX  
 OS Piscirickettsia salmonis.  
 OS Synthetic.  
 OS  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 1..483  
 FT /tag= a  
 FT /partial= "Ospa"  
 FT /product= "Genus specific 17kDa antigen, the sequence does  
 FT /note= "not include a stop codon"  
 PN  
 PN CA2281913-A1.  
 PD 17-MAR-2001.  
 XX  
 PD 17-SEP-1999; 99CA-2281913.  
 PF 17-SEP-1999; 99CA-2281913.  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 XX (KAYW/) KAY W. W.  
 PA (BURI/) BURIAN J.  
 PA (KUZV/) KUZV M. A.  
 XX  
 PI Kay WW, Burian J, Kuzk MA;  
 XX  
 DR MPI: 2001-316844/34.  
 DR P-PSDB: AAB81127.  
 XX  
 PT Method for protecting poikilothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis -  
 XX  
 PS Example 3; Fig 4C; 35pp; English.  
 XX  
 .CC This invention relates to a method for the protection against infection  
 .CC of a poikilothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,

CC particularly polkilothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents DNA which has been optimised for expression in  
 CC Escherichia coli to encode the P. salmonis Ospa protein. An Ospa protein  
 CC with an N-terminal fusion partner is used in a vaccine to create an  
 CC anti-Ospa antibody response.

SO Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 62.9%; Score 483; DB 22; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 2,6e-131;

Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 ATGCGTGGTTCCTGCGAGGCGACCTCTGATATATATCTCTTTTCTCGTGGGTTGC 345  
 DB 1 ATGCGGTGGTTCCTGCGAGGCGACCTCTGATATATATCTCTTTTCTCGTGGGTTGC 60  
 OY 346 GCCAGAACTTTCAGCCGCGAGAAAGTTGGGCGGCGCCAGCGGTGGTGGGCGGTGT 405  
 DB 61 GCCAGAACTTTCAGCCGCGAGAAAGTTGGGCGGCGCCAGCGGTGGTGGGCGGTGT 120  
 OY 406 GCCGCGACCTGTTCGTAAGGCTGTGTCGTGTCGATGGCATTGCGGCGGTG 465  
 DB 121 GCCGCGACCTGTTCGTAAGGCTGTGTCGTGTCGATGGCATTGCGGCGGTG 180  
 OY 466 CTGGGCGGTCTGATTTGGCTCTTAATTCGGTCAGACATGAGCAGCAGAGTAAATCAA 525  
 DB 181 CTGGGCGGTCTGATTTGGCTCTTAATTCGGTCAGACATGAGCAGCAGAGTAAATCAA 240  
 OY 526 CTGAACCACTCTCTGAAAAAGTGAAGCGGCGCAGGTTACTGTTGGCGTAAATCCGGAC 585  
 DB 241 CTGAACCACTCTCTGAAAAAGTGAAGCGGCGCAGGTTACTGTTGGCGTAAATCCGGAC 300  
 OY 586 ACCGTAACAGACTCTCTGTGGAACCGGTTGCGACCTACCGCGTTACAAACAAGAGAA 645  
 DB 301 ACCGTAACAGACTCTCTGTGGAACCGGTTGCGACCTACCGCGTTACAAACAAGAGAA 360  
 OY 646 CCGCGTACAGACTCTCTGTGGAACCGGTTGCGACCTACCGCGTTACAAACAAGAGAA 705  
 DB 361 CCGCGTACAGACTCTCTGTGGAACCGGTTGCGACCTACCGCGTTACAAACAAGAGAA 420  
 OY 706 GAAATCTACGCGACCGCGTGCCTCAGCCGATGCGCGTGGCAGGTGATTACACCGAA 765  
 DB 421 GAAATCTACGCGACCGCGTGCCTCAGCCGATGCGCGTGGCAGGTGATTACACCGAA 480  
 OY 766 AAA 768  
 DB 481 AAA 483

RESULT 3  
 AAF86246  
 ID AAF86246 standard; DNA; 486 BP.

XX AAF86246;  
 AC 11-JUL-2001 (first entry)  
 XX Ospa 17kD antigen gene.  
 DE Polkilothermic fish: Piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;  
 KM SRS; ds.  
 XX Piscirickettsia salmonis.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..486  
 FT /tag- a  
 FT /partial  
 FT /product= "Ospa"  
 FT /note= "Genus specific 17kDa antigen, the sequence does

FT not include a stop codon"

XX CA2281913-A1.

XX 17-MAR-2001.

XX 17-SEP-1999; 99CA-2281913.

XX 17-SEP-1999; 99CA-2281913.

XX (KAYW/) KAY W W.

XX (BURI/) BURIAN J.

XX (KUZK/) KUZK M A.

XX Kay WW, Burian J, Kuzk MA;

XX WPI: 2001-316844/34.

XX P-PSDB: AAB81126.

PT Method for protecting polkilothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis

CC This invention relates to a method for the protection against infection  
 CC of a polkilothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkilothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents DNA encoding the P. salmonis Ospa protein. An Ospa  
 CC protein with an N-terminal fusion partner is used in a vaccine to create  
 CC an anti-Ospa antibody response.

SO Sequence 486 BP; 137 A; 79 C; 144 G; 126 T; 0 other;

Query Match 38.4%; Score 295; DB 22; Length 486;  
 Best Local Similarity 76.0%; Pred. No. 3,1e-76;

Matches 364; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

OY 290 GTGGTTCGCGAGGCGACCTCTGATCATTTCTGTTTCCGTGGGTTGGCGCC 349  
 DB 8 GAGGATGTTTCAAGGAGTAGTCTATTTATTCAGTGTGTTTAACTGCTGCGCC 67  
 OY 350 AGAATTTACAGCCCGCAGAGAGTTGGCGCGCCACCGGTGGTGGGCGGTGTCGG 409  
 DB 68 AGAATTTACAGCCCGCAGAGAGTTGGCGCGCCACCGGTGGTGGGCGGTGTCGG 127  
 OY 410 GCCAGCTGTGCTGTAAGGCTGTGTCGTGTCGATGSCCATGCGCGGTCTGG 469  
 DB 128 GCCAGCTGTGCTGTAAGGCTGTGTCGTGTCGATGSCCATGCGCGGTCTGG 187  
 OY 470 GCGGTGATGTCCTGTAAGGCTGTGTCGTGTCGATGSCCATGCGCGGTCTGG 529  
 DB 188 GTGGATTAATTTGTTCTTAATATCGTCAATGATGTCAGCAGGATTAATTAAGCTAA 247  
 OY 530 ACCAGTCTGGAAGAAAGTGAAGCCGCGCAGGTACTGCTGCGCTAATCCGAGACCG 589  
 DB 248 ACCAGTCTGGAAGAAAGTGAAGCCGCGCAGGTACTGCTGCGCTAATCCGAGATTCAG 307  
 OY 590 GTAAACAGTACTCTGTGGAACCGGTTGCGACCTACACGCTTACAAACAAGAGACGCC 649  
 DB 308 GCAATAGTATTACTGTGACACCGTGTGCTTACACGCTTACAAACAAGAGACGCC 367  
 OY 650 GTACAGAGTACTGCGGATTCAGAGAAAGCAATGATGCGAGTCAAGACAGAA 709  
 DB 368 GCCAGCAATATTTGTGAGAAATTTTACGAAAAAGCGATGATGAGGCGCAACCAAGAGA 427  
 OY 710 TCTACGCGACCGGCGCTCAGCCGAGTGGCGGTGCGAGGTGATTACCGCAAAAA 768

Db	188	GTGATTTAATTTGGTTCTTAAATAATCGTCATCGATGTGATCAGCAGATGAAATTAACCTTAA	247
Oy	530	ACCAAGTCTCGGAAAAAGTAAAGCGCGCAGAGTTACTGCTTGGCCGTAAATCCGACACCG	589
Db	248	ACCAAGTATTGGAAAGGTAAAGCAAGGCAAGTACACGTTGGCGTAATCCAGATACAG	307
Oy	590	GTAACACCTCTCTGTGTGGAACCGGTTTCCACCTTACACAGCGTTACACAAACAGGAAGCC	649
Db	308	GCAATGATTTTACTGTTGGCCAGTGCCTACTTACACAGCGTTTACAATTAAGCAAGACGCTC	367
Oy	650	GTCAGCAGTACTGCCCCGATTTTCAGCAGAAAGCCATGATCGCAGGTCAAGAAACAGGAA	709
Db	368	GCCAGCAATATTGTGAGATATTTTCAGCAAAAAGGCGATGATTGCAGGCGACAGACAGAG	427
Oy	710	TCTACGGCACCCTGGTCCCTTACGCGGATGCGCGCTTGGCAGGTGATTTAGCACCCGAAAA	768
Db	428	TTTACGGCACTGCATGCGGCAACCGGATGTGCTTGGCAAGTCATTTCAACAGCAAAAA	486
RESULT 5			
AAD11043			
ID	AAD11043 standard; DNA; 573 BP.		
XX	AAD11043;		
XX	24-SEP-2001 (first entry)		
DE	Clostridium cellulovorans cellulose binding domain-180 (CBD-180) DNA.		
XX			
XX	Polysaccharide modification; polysaccharide binding domain; PBD; paper		
KW	yarn; fiber; textiles; biological crosslinker; mechanical property;		
KW	Clostridium cellulovorans cellulose binding domain; wet strength;		
KW	durability; elasticity; CBDclos; cellulose binding protein A; CBP A;		
XX	CBD-180; ds.		
XX	Clostridium cellulovorans.		
OS			
XX			
FH	Key		
FT	location/Qualifiers		
FT	3..572		
FT	/*tag= a		
FT	/product= "Cellulose binding domain-180 protein"		
FT	/note= "CDS does not include stop codon"		
XX	/partial		
PN	MO200134091-A2.		
XX			
PD	17-MAY-2001.		
XX			
PF	02-NOV-2000; 2000MO-IL00708.		
XX			
PR	08-NOV-1999; 99US-0164140.		
PR	18-NOV-1999; 99US-0166389.		
XX			
PA	(CBDT-) CBD TECHNOLOGIES LTD.		
PA	(VLISS ) YISSOM RES DEV CO HEBREW UNIV JERUSALEM.		
PI	Levy I, Nussinovitch A, Shoseyov O;		
XX			
DR	WPI: 2001-457121/49.		
XX	P-PSDB: AAEO5746.		
PT	Preparation of a polysaccharide containing material having at least one		
PT	desired structural, chemical, physical, electrical and/or mechanical		
XX	property		
XX			
PS	Example 1.2: Fig 1e-1g; 121pp; English.		
XX			
CC	The present invention relates to methods and compositions for cross-		
CC	linking and/or modifying the properties of polysaccharide materials.		
CC	The method involves treating the polysaccharide structure with a		
CC	polysaccharide binding domain (PBD) fusion protein. The method is		
CC	used to alter the structural, chemical, physical, electrical and		

CC mechanical properties of polysaccharide materials such as paper,  
CC yarns, fibers and textiles, using biological crosslinking agents.  
CC The polysaccharide containing materials have improved mechanical  
CC properties such as wet strengths, durability and elasticity. The PBD  
CC reagent is applied in the forming stage in fluting paper manufacture  
CC which eliminates the sizing step. The use of a biological crosslinker  
CC improves the recyclability of paper products. The PBD reagent maintains  
CC the fine fibers in a slurry therefore resulting in better recovery of  
CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
CC which enhances the ability of the alkaline glue used in binding to  
CC penetrate paper.  
CC The present sequence is a DNA encoding Clostridium cellulovorans  
CC cellulose binding domain (CBD)-180. CBD-180 contains 180 amino acids of  
CC Clostridium cellulovorans CBD (CBDClos) of cellulose binding protein A  
CC (CBP A).  
CC  
XX  
SQ Sequence 573 BP; 215 A; 112 C; 98 G; 148 T; 0 other;

Query Match 35.4%; Score 271.8; DB 22; Length 573;  
Best Local Similarity 99.3%; Pred. No. 2.1e-69;  
Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCATTA 60  
DB 3 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCATTA 62  
QY 61 ATCAAAATTTACTACACATCTGACAGATTTAAATTTAAATGACGTAAGTATAT 120  
DB 63 ATCAAAATTTACTACACATCTGACAGATTTAAATTTAAATGACGTAAGTATAT 122  
QY 121 TATTACACAAGTGATGTGATACAGCAAACTTCTGCTGACCAATGCTGTCATTA 180  
DB 123 TATTACACAAGTGATGTGATACAGCAAACTTCTGCTGACCAATGCTGTCATTA 182  
QY 181 TTAGGAATAGCTATGTTGATACACAGCAAGCAAACTTCTGTTAAAGAAACA 240  
DB 183 TTAGGAATAGCTATGTTGATACACAGCAAGCAAACTTCTGTTAAAGAAACA 242  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGCA 275  
DB 243 GCAAGCCCAACATCAACCTATGATACATATCTGCA 277

RESULT 6  
AAD11044  
ID AAD11044 standard; DNA: 1030 BP.  
XX  
AC AAD11044:  
XX  
DT 24-SEP-2001 (first entry)  
XX  
DE Clostridium cellulovorans CBD cross linker protein (CCP) encoding DNA.  
XX  
KM Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
KM yarn; fiber; textile; biological crosslinker; mechanical property;  
KM wet strength; durability; elasticity; cellulose binding domain; CBD;  
KM CBD cross linker protein; CCP; ds.  
XX  
OS Clostridium cellulovorans.  
XX  
FH Key Location/Qualifiers  
FT CDS 3..1028  
FT /tag= a  
FT /product= "CCP protein"  
FT /transl\_except= "(pos: 1020..1028, aa:Pro-Asp)"  
FT /note= "CDS does not include stop codon"  
XX  
XX /partial  
XX  
PN MO200134091-A2.  
XX  
XX 17-MAY-2001.  
XX  
PF 02-NOV-2000; 2000MO-IL00708.

XX  
PR 08-NOV-1999; 99US-0164140.  
PR 18-NOV-1999; 99US-0166389.  
XX  
PA (CBDT-) CBD TECHNOLOGIES LTD.  
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
XX  
PI Levy I, Nussimovitch A, Shoseyov O;  
XX  
XX WPI; 2001-457121/49.  
DR P-PSDB; AAE05747.  
XX  
PT Preparation of a polysaccharide containing material having at least one  
PT desired structural, chemical, physical, electrical and/or mechanical  
PT property -  
XX  
PS Example 1.2; Fig 2b-2e; 121pp; English.  
XX  
CC The present invention relates to methods and compositions for cross-  
CC linking and/or modifying the properties of polysaccharide materials.  
CC The method involves treating the polysaccharide structure with a  
CC polysaccharide binding domain (PBD) fusion protein. The method is  
CC used to alter the structural, chemical, physical, electrical and  
CC mechanical properties of polysaccharide materials such as paper,  
CC yarns, fibers and textiles, using biological crosslinking agents.  
CC The polysaccharide containing materials have improved mechanical  
CC properties such as wet strengths, durability and elasticity. The PBD  
CC reagent is applied in the forming stage in fluting paper manufacture  
CC which eliminates the sizing step. The use of a biological crosslinker  
CC improves the recyclability of paper products. The PBD reagent maintains  
CC the fine fibers in a slurry therefore resulting in better recovery of  
CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
CC which enhances the ability of the alkaline glue used in binding to  
CC penetrate paper.  
CC The present sequence is a DNA encoding Clostridium cellulovorans  
CC cellulose binding domain (CBD) cross linker protein (CCP).  
CC  
XX  
SQ Sequence 1030 BP; 402 A; 195 C; 166 G; 267 T; 0 other;

Query Match 35.4%; Score 271.8; DB 22; Length 1030;  
Best Local Similarity 99.3%; Pred. No. 2.9e-69;  
Matches 273; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCATTA 60  
DB 3 ATGTCAGTGAATTCACAACTCTAACAAATCAGCACAACAACTCAATTACACCATTA 62  
QY 61 ATCAAAATTTACTACACATCTGACAGATTTAAATTTAAATGACGTAAGTATAT 120  
DB 63 ATCAAAATTTACTACACATCTGACAGATTTAAATTTAAATGACGTAAGTATAT 122  
QY 121 TATTACACAAGTGATGTGATACAGCAAACTTCTGCTGACCAATGCTGTCATTA 180  
DB 123 TATTACACAAGTGATGTGATACAGCAAACTTCTGCTGACCAATGCTGTCATTA 182  
QY 181 TTAGGAATAGCTATGTTGATACACAGCAAGCAAACTTCTGTTAAAGAAACA 240  
DB 183 TTAGGAATAGCTATGTTGATACACAGCAAGCAAACTTCTGTTAAAGAAACA 242  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGCA 275  
DB 243 GCAAGCCCAACATCAACCTATGATACATATCTGCA 277

RESULT 7  
AAQ72917  
ID AAQ72917 standard; DNA: 486 BP.  
XX  
AC AAQ72917:  
XX  
DT 05-JUN-1995 (first entry)  
XX  
DE Cellulose binding domain.

Query Match	35.2%	Score 270.2	DB 15	Length 486
Best Local Similarity	98.9%	Pred. No. 5,76	69	
Matches 272	Conservative 0	Mismatches 3	Indels 0	Gaps
QY	1	ATGTCAGTTGAATTTCTACACCTCTTACAAATCAGCACACAAACAACTCATATTCACCATTA	60	
Db	16	ATGCAATCTGGAATTTTATACAACTCTTACAAATCAGCACAAACAAACAACTCATATTCACCATTA	75	
QY	61	ATCAAAATTACTACACACATCTGACAGTGATTTAAATTTAAAGAGCTAAAGTTGATAT	120	
Db	76	ATCAAAATTACTACACACATCTGACAGTGATTTAAATTTAAAGAGCTAAAGTTGATAT	135	
QY	121	TATTACCAAGTAGTAGTGCACAAAGGACAACTTCTGGTGTGACCATGCTGGTCATTA	180	
Db	136	TATTACCAAGTAGTAGTGCACAAAGGACAACTTCTGGTGTGACCATGCTGGTCATTA	195	
QY	181	TTAGGAATATGCTATGTTGATATACATACAGAAAGTGACAGCAAACTCTGGTTAAAGAACCA	240	
Db	196	TTAGGAATATGCTATGTTGATATACATACAGAAAGTGACAGCAAACTCTGGTTAAAGAACCA	255	
QY	241	GCAAGCCCAACATCAACTATGATATATCTGGA	275	
Db	256	GCAAGCCCAACATCAACTATGATATATCTGGA	290	

2

DB 136 TATTACACAAGTGTGATGACACAGCAAACTTCTGCTGTGACCATGCTGGTCATTTA 195  
 QY 181 TTAGSAAATAGCTATGTGTGATACTAGCAAGTGCAGCAACTTCGTTAAGAACA 240  
 DB 196 TTAGSAAATAGCTATGTGTGATACTAGCAAGTGCAGCAACTTCGTTAAGAACA 255  
 QY 241 GCAGCCCAACATCAACCTATGATACATATCTGGA 275  
 DB 256 GCAGCCCAACATCAACCTATGATACATATGTTGA 290

RESULT 9  
 AAX24930  
 ID AAX24930 standard; DNA; 499 BP.  
 XX AAX24930;  
 AC  
 XX  
 XX 21-JUN-1999 (first entry)  
 DT  
 XX  
 XX Clostridium cellulovorans cbpa cellulose binding domain DNA.  
 DE  
 XX Cellulose binding domain; CBD; cbpa; endo-1,4-beta-glucanase; Cell;  
 KW Arabidopsis thaliana; transgenic plant; crop improvement;  
 KM morphology; cell wall; ds.  
 XX  
 OS Clostridium cellulovorans.  
 PN WO9907830-A1.  
 XX 18-FEB-1999.  
 PD  
 XX 26-JUL-1998; 98WO-IL00345.  
 PF  
 XX 13-JAN-1998; 98US-0006636.  
 PR 27-JUL-1997; 97IL-0121404.  
 PR 13-JAN-1998; 98US-0006632.  
 XX  
 XX (YISS ) YISSUM RES & DEV CO.  
 PA  
 XX Shani Z, Shoseyov O, Shplegl E;  
 PI WPI: 1999-180488/15.  
 DR  
 XX Transgenic plants expressing cell-wall modulating protein - have  
 PT altered morphology, e.g. Increased growth, modified fiber length or  
 PT cellulose content  
 PS  
 XX Disclosure; Page 134; 144pp; English.  
 PS  
 XX This DNA fragment encodes the cellulose binding domain (CBD) of  
 CC the cbpa protein of Clostridium cellulovorans. It was obtained  
 CC by PCR amplification (see also AAX24932-53), and was used in the  
 CC construction of binary vector pCC1, in which cbd was joined to  
 CC the promoter and signal region of the novel endo-1,4-beta-glucanase  
 CC cell gene (see AAX24923) of Arabidopsis thaliana. Expression of cbd  
 CC in transgenic tobacco plants modulated their growth. The cell  
 CC promoter can provide expression of any protein in elongating  
 CC tissue. CBD is an example of a cell wall modulation transgene  
 CC used to alter the structure or morphology of a plant. Transgenic  
 CC plants of the invention may have altered biomass, growth, yield,  
 CC greater or less resistance to biodegradation, be more or less  
 CC digestible by ruminants, have altered cellulose content, larger or  
 CC smaller leaves, etc., when compared to non-transgenic plants of the  
 CC same species.  
 CC  
 SQ Sequence 499 BP; 197 A; 93 C; 82 G; 127 T; 0 other;  
 XX  
 XX  
 Query Match 35.2%; Score 270.2; DB 20; Length 499;  
 Best Local Similarity 98.9%; Pred. No. 5.8e-69;  
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 ATGTCAAGTTGAATTTTACAACTCTAACAAATCGACAAACAACTCAATTACACCAATA 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 21 ATGTCAAGTTGAATTTTACAACTCTAACAAATCGACAAACAACTCAATTACACCAATA 80  
 QY 61 ATCAAAATTTACTTAACACATCTGACAGTGAATTTAAATTAATGCTAAAGTAACTAT 120  
 DB 81 ATCAAAATTTACTTAACACATCTGACAGTGAATTTAAATTAATGCTAAAGTAACTAT 140  
 QY 121 TATTACACAAGTGTGATGATGACACAGCAAACTTCTGCTGTGACCATGCTGGTCATTTA 180  
 DB 141 TATTACACAAGTGTGATGATGACACAGCAAACTTCTGCTGTGACCATGCTGGTCATTTA 200  
 QY 181 TTAGSAAATAGCTATGTGTGATACTAGCAAGTGCAGCAAACTTCGTTAAGAACA 240  
 DB 201 TTAGSAAATAGCTATGTGTGATACTAGCAAGTGCAGCAAACTTCGTTAAGAACA 260  
 QY 241 GCAGCCCAACATCAACCTATGATACATATCTGGA 275  
 DB 261 GCAGCCCAACATCAACCTATGATACATATGTTGA 295

RESULT 10  
 AAD11042  
 ID AAD11042 standard; DNA; 507 BP.  
 XX AAD11042;  
 AC  
 XX 24-SEP-2001 (first entry)  
 DT  
 XX  
 XX Clostridium cellulovorans cellulose binding domain (CBDclos) DNA.  
 DE  
 XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
 KW yaro; fiber; textile; biological crosslinker; mechanical property;  
 KM Clostridium cellulovorans cellulose binding domain; wet strength;  
 KM durability; elasticity; CBDclos; cellulose binding protein A; CBP A; ds.  
 XX  
 OS Clostridium cellulovorans.  
 PN WO200134091-A2.  
 XX 17-MAY-2001.  
 PD  
 XX 02-NOV-2000; 2000WO-IL00708.  
 PF  
 XX 08-NOV-1999; 99US-0164140.  
 PR 18-NOV-1999; 99US-0166389.  
 XX  
 XX (CBDT-) CBD TECHNOLOGIES LTD.  
 PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
 PI Levy I, Nussinovitch A, Shoseyov O;  
 XX WPI: 2001-457121/49.  
 DR P-PsDB: AAE05745.  
 DR  
 XX Preparation of a polysaccharide containing material having at least one  
 PT desired structural, chemical, physical, electrical and/or mechanical  
 PT property -  
 PS Example 1.1; Page 111; 121pp; English.  
 PS  
 XX The present invention relates to methods and compositions for cross-  
 CC linking and/or modifying the properties of polysaccharide materials.  
 CC The method involves treating the polysaccharide structure with a  
 CC polysaccharide binding domain (PBD) fusion protein. The method is  
 CC used to alter the structural, chemical, physical, electrical and  
 CC mechanical properties of polysaccharide materials such as paper,  
 CC yarns, fibers and textiles, using biological crosslinking agents.  
 CC The polysaccharide containing materials have improved mechanical  
 CC properties such as wet strengths, durability and elasticity. The PBD

CC reagent is applied in the forming stage in fluting paper manufacture  
 CC which eliminates the sining step. The use of a biological crosslinker  
 CC improves the recyclability of paper products. The PBD reagent maintains  
 CC the fine fibers in a slurry therefore resulting in better recovery of  
 CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
 CC which enhances the ability of the alkaline glue used in binding to  
 CC penetrate paper.  
 CC The present sequence is a DNA encoding Clostridium cellulovorans  
 CC cellulose binding domain (CBdClos) of cellulose binding protein A  
 CC (CBP A).  
 CC  
 XX Sequence 507 BP; 200 A; 94 C; 84 G; 129 T; 0 other;

Query Match Best Local Similarity 35.2%; Score 270.2; DB 22; Length 507;  
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGAATTTCTACACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 60  
 DB 21 ATGTCAGTTGAATTTCTACACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 80  
 OY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATGACGTAAGTATATAT 120  
 DB 81 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATGACGTAAGTATATAT 140  
 OY 121 TATTACACAAGTGATGTACACAAGCAAACTTTCTGGTGTGACCATGCTGTGATTA 180  
 DB 141 TATTACACAAGTGATGTACACAAGCAAACTTTCTGGTGTGACCATGCTGTGATTA 200  
 OY 181 TTAGGAATAGCTATGCTGTGTAACACTAGCAAGTACAGCAAACTGCTTTAAAGAA 240  
 DB 201 TTAGGAATAGCTATGCTGTGTAACACTAGCAAGTACAGCAAACTGCTTTAAAGAA 260  
 OY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275  
 DB 261 GCAAGCCCAACATCAACCTATGATATATCTGGA 295

RESULT 11  
 AAD11046  
 ID AAD11046 standard; DNA; 984 BP.

AC AAD11046;  
 XX  
 DT 24-SEP-2001 (first entry)

DE Chimeric S peptide-cellulose binding domain-S protein encoding DNA.  
 XX  
 KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
 KW yarn; fiber; textile; biological crosslinker; mechanical property;  
 KW wet strength; durability; elasticity; cellulose binding domain; CBD;  
 KW chimeric protein; S peptide-cellulose binding domain-S protein;  
 KW Sppe-CBD-Sprot; bovine; ds.

OS Chimeric - Clostridium cellulovorans.  
 OS Chimeric - Bos sp.

XX  
 Key Location/Qualifiers  
 FT 1 984  
 FT CDS

FT /tag= a  
 FT /product= "Sppe-CBD-Sprot protein"  
 FT /transl\_except= "(pos: 979..981, aa:Xaa)"  
 FT /note= "Xaa corresponds to in-frame stop codon;"  
 FT CDS does not include start and stop codon."  
 FT /partial  
 FT 68..624  
 FT /tag= b  
 FT /note= "This region is derived from C. cellulovorans"  
 FT misc\_feature  
 FT 652..981  
 FT /tag= c  
 FT /note= "This region is derived from bovine"  
 XX  
 PN WO200134091-A2.

XX  
 PD 17-MAY-2001.  
 XX  
 PF 02-NOV-2000; 2000WO-IL00708.  
 XX  
 PR 08-NOV-1999; 99US-0164140.  
 PR 18-NOV-1999; 99US-0166389.  
 XX  
 PA (CBPT-) CBD TECHNOLOGIES LTD.  
 PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
 XX  
 PI Levy I, Nussinovitch A, Shoseyov O;  
 DR WPI; 2001-457121/49.  
 DR P-PSDB; AAE05749.

Preparation of a polysaccharide containing material having at least one  
 desired structural, chemical, physical, electrical and/or mechanical  
 property

Example 4; Fig 4b-4g; 121pp; English.

The present invention relates to methods and compositions for cross-  
 linking and/or modifying the properties of polysaccharide materials.  
 The method involves treating the polysaccharide structure with a  
 polysaccharide binding domain (PBD) fusion protein. The method is  
 used to alter the structural, chemical, physical, electrical and  
 mechanical properties of polysaccharide materials such as paper,  
 yarns, fibers and textiles, using biological crosslinking agents.  
 The polysaccharide containing materials have improved mechanical  
 properties such as wet strengths, durability and elasticity. The PBD  
 reagent is applied in the forming stage in fluting paper manufacture  
 which eliminates the sining step. The use of a biological crosslinker  
 improves the recyclability of paper products. The PBD reagent maintains  
 the fine fibers in a slurry therefore resulting in better recovery of  
 raw materials. The PBD molecules are eluted by strong alkaline conditions  
 which enhances the ability of the alkaline glue used in binding to  
 penetrate paper.  
 The present sequence is a DNA encoding S peptide-cellulose binding  
 domain-S protein (Sppe-CBD-Sprot), a fusion protein derived from  
 Clostridium cellulovorans and bovine.

Sequence 984 BP; 325 A; 240 C; 193 G; 226 T; 0 other;

Query Match Best Local Similarity 35.2%; Score 270.2; DB 22; Length 984;  
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATGTCAGTTGAATTTCTACACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 60  
 DB 103 ATGTCAGTTGAATTTCTACACTCTAACAAATCAGCACAACAACTCAATTTACACCAATA 162  
 OY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATGACGTAAGTATATAT 120  
 DB 163 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATGACGTAAGTATATAT 222  
 OY 121 TATTACACAAGTGATGTACACAAGCAAACTTTCTGGTGTGACCATGCTGTGATTA 180  
 DB 223 TATTACACAAGTGATGTACACAAGCAAACTTTCTGGTGTGACCATGCTGTGATTA 282  
 OY 181 TTAGGAATAGCTATGCTGTGTAACACTAGCAAGTACAGCAAACTGCTTTAAAGAA 240  
 DB 283 TTAGGAATAGCTATGCTGTGTAACACTAGCAAGTACAGCAAACTGCTTTAAAGAA 342  
 OY 241 GCAAGCCCAACATCAACCTATGATATATCTGGA 275  
 DB 343 GCAAGCCCAACATCAACCTATGATATATCTGGA 377

RESULT 12  
 AA227550  
 ID AA227550 standard; DNA; 1146 BP.  
 XX



AC AA227550:  
 XX  
 XX 13-DEC-1999 (first entry)  
 DT  
 XX  
 DE Gausia Luciferase fusion protein coding sequence.  
 XX  
 XX Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;  
 KM bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;  
 KM body paint; squirt gun; balloon; slimy play material; soap; toothpaste;  
 KM fusion protein; ds.  
 XX  
 XX Gausia sp.  
 OS  
 XX  
 XX MO949019-A2.  
 PN  
 XX 30-SEP-1999.  
 PD  
 XX  
 XX 26-MAR-1999; 99WO-US06698.  
 PF  
 XX  
 XX 27-MAR-1998; 98US-0079624.  
 PR 15-JUN-1998; 98US-0089367.  
 PR 01-OCT-1998; 98US-0102939.  
 XX  
 PA (PROL-) PROLUME LTD.  
 PA (BRYA/) BRYAN B J.  
 XX  
 PI Bryan BJ, Szent-Gyorgyi C;  
 XX  
 XX WPI; 1999-580443/49.  
 DR P-PSDB; AAY39952.  
 DR  
 XX  
 XX  
 PT New isolated Renilla mulleri, Gausia and Pleuromamma Luciferase and  
 PT Renilla and Ptilosarcus green fluorescent protein nucleic acids -  
 XX  
 XX  
 PS Disclosure; Page 222-223; 233pp; English.  
 XX  
 XX  
 CC This sequence encodes a Luciferase of the invention. The invention  
 CC relates to Renilla mulleri, Gausia and Pleuromamma Luciferase and  
 CC Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and  
 CC proteins. The Luciferases and GFPs can be used in  
 CC bioluminescence-generating systems, assays, screening methods, diagnostic  
 CC method and articles of manufacture. They can be expressed using  
 CC e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla  
 CC mulleri, Gausia and Pleuromamma Luciferase or Renilla or Ptilosarcus  
 CC GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,  
 CC fairy dust, beverages, soft drinks, foods, textile products, bubbles,  
 CC balloons, personal items, dentifrices, soaps, body paints, bubble bath,  
 CC ink or paper products. In particular, they can be used in e.g. squirt  
 CC guns, pellet guns, finger paints, foot bags, greeting cards, slimy play  
 CC material, clothing, bubble making toys, bath powders, cosmetics, body  
 CC lotions, gels, body powders, toothpastes, mouthwashes,  
 CC soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,  
 CC frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,  
 CC ice, dry ice or fountains. The nucleic acids can also be used to produce  
 CC transgenic fish and plants.  
 CC  
 XX  
 XX Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;

Query Match 35.2%; Score 270.2; DB 20; Length 1146;  
 Best Local Similarity 98.9%; Pred. No. 9e-69;  
 Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAAGTGAATTCACACTCTAACCAATTCAGCAACAAACCTCAATTACACCATTA 60  
 Db 1 ATGTCAAGTGAATTCACACTCTAACCAATTCAGCAACAAACCTCAATTACACCATTA 60  
 QY 61 ATCAAAATTTACTAACACATCTGACAGATTTAAATTAATGACGCTAAAGTTAGTAT 120  
 Db 61 ATCAAAATTTACTAACACATCTGACAGATTTAAATTAATGACGCTAAAGTTAGTAT 120  
 QY 121 TATTACACAAGTGTGTACACAAGCAAACTTCTGTGTGTGACCAATGCTGTGCATTA 180  
 Db 121 TATTACACAAGTGTGTACACAAGCAAACTTCTGTGTGTGACCAATGCTGTGCATTA 180

QY 181 TTAGGAATAGCTATGTGTGTAACACTAGCAAGTGCACGCAAACTTCGTTAAGAACA 240  
 Db 181 TTAGGAATAGCTATGTGTGTAACACTAGCAAGTGCACGCAAACTTCGTTAAGAACA 240  
 QY 241 GCAAGCCCAACATCAACCTGTGATACATATCTGGA 275  
 Db 241 GCAAGCCCAACATCAACCTGTGATACATATGTTGA 275  
 RESULT 13  
 AAD22201  
 ID AAD22201 standard; DNA; 1146 BP.  
 XX  
 AC AAD22201;  
 XX  
 DT 12-FEB-2002 (first entry)  
 XX  
 DE Gausia species CBD-Luciferase fusion protein encoding DNA.  
 XX  
 KW Green fluorescent protein; GFP; bioluminescence generating system; toy;  
 KW Luciferase; finger paint; slimy play material; fishing lure; sparkler;  
 KW doll; balloon; personal care item; cosmetic; bath powder; body cream;  
 KW tooth paste; mouth wash; soap; body paint; ornamental transgenic plant;  
 KW bubble bath; fountain; fairy dust; food; gelatin; icing; frosting; beer;  
 KW wine; champagne; milk; soft drink; ice cube; ice; foot bag; beverage;  
 KW dry ice; textile; clothing; paper product; greeting card; wrapping paper;  
 KW fusion protein; ds.  
 XX  
 XX Gausia sp.  
 OS  
 XX  
 XX  
 FH Location/Qualifiers  
 FT 1..1146  
 FT CDS /tag= a  
 FT /product= "Gausia CBD-Luciferase fusion protein"  
 FT /note= "CDS does not include start and stop codon"  
 FT /partial  
 XX  
 PN MO200168824-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 15-MAR-2001; 2001WO-US08277.  
 XX  
 PR 15-MAR-2000; 2000US-189691P.  
 XX  
 PA (PROL-) PROLUME LTD.  
 PA (BRYA/) BRYAN B J.  
 XX  
 PI Bryan BJ, Szent-Gyorgyi C, Szczepaniak W;  
 XX  
 XX WPI; 2002-010561/01.  
 DR P-PSDB; AAE13383.  
 XX  
 PT Nucleic acids encoding Renilla reniformis green fluorescent proteins,  
 PT useful in diagnostic bioluminescence procedures -  
 XX  
 XX  
 PS Disclosure; Page 162-163; 175pp; English.  
 XX  
 XX The patent discloses sea pansy (Renilla reniformis) green fluorescent  
 CC proteins (GFP) and their corresponding polynucleotides. The invention  
 CC also relates to sequences of the bioluminescence generating system  
 CC (e.g. Luciferase). R. reniformis GFP are used in diagnostic methods  
 CC and in the production of novelty items such as toys (e.g. squirt gun,  
 CC pellet gun, toy "Halloween" eggs, toy cigarettes, board/card game  
 CC toys), finger paints, slimy play material, bubbles in bubble making  
 CC toys, fishing lures, dolls, sparklers, magic wand toys, balloons,  
 CC personal care item (e.g. cosmetic, bath powders, body creams, tooth  
 CC pastes, mouth wash, soaps, body paints, bubble bath), ornamental  
 CC transgenic plants, fountain, fairy dust, food (gelatins, icings,  
 CC frosting, beers, wine, champagne, milk, soft drinks, ice cubes, ice,  
 CC dry ice, beverage), textile (foot bag, clothing) and/or paper product  
 CC (greeting cards, wrapping paper). The present sequence is a DNA



```

OS  Piscirickettsia salmonis.
XX
PN  CA2281913-A1.
XX
PD  17-MAR-2001.
XX
PF  17-SEP-1999; 99CA-2281913.
XX
PR  17-SEP-1999; 99CA-2281913.
XX
PA  (KAYM/) KAY W W.
PA  (BURI/) BURIAN J.
XX  (KUZK/) KUZK M A.
XX
PI  Kay WW, Burian J, Kuzk MA;
XX
DR  WPI: 2001-316844/34.
XX
PT  Method for protecting polkilothermic fish against salmonid rickettsial
PT  septicemia and other rickettsial diseases comprises administering a
PT  vaccine containing the OSPA protein of Piscirickettsia salmonis -
XX
XX
PS  Example 3; Fig 4B; 35pp; English.
XX
XX  This invention relates to a method for the protection against infection
XX  of a polkilothermic fish by the bacterial pathogen, Piscirickettsia
XX  salmonis. The method comprises administering an immunogenic amount of a
XX  P. salmonis specific antigen termed OSPA, or an immunogenic fragment of
XX  OSPA in the form of a vaccine. The method is used for protecting animals,
XX  particularly polkilothermic fish, against the bacterial pathogen
XX  P. salmonis. The method is also useful for protecting against salmonid
XX  rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX  sequence represents a PCR primer used in the cloning and optimisation of
XX  the P. salmonis OSPA gene. The OSPA gene is used in the method of the
XX  invention.
XX
SQ  Sequence 118 BP: 21 A; 32 C; 28 G; 37 T; 0 other:

```

## Query Match

Best Local Similarity 15.4%; Score 118; DB 22; Length 118;  
 Pred. No. 1e-24; Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY  493 GGTGAGAGCATGAGACAGACGATATAATCAAACTGAACAGCTCTGGAAAAAGTGAA 552
    |||||||
DB  118 GGTGAGAGCATGAGACAGACGATATAATCAAACTGAACAGCTCTGGAAAAAGTGAA 59
    |||||||
QY  553 GCCGGCCAGGTTACTGTTGGCGTAATCCGGACACCGGTAACAGCTACTGTGTGGAAC 610
    |||||||
DB  58 GCCGGCCAGGTTACTGTTGGCGTAATCCGGACACCGGTAACAGCTACTGTGTGGAAC 1
    |||||||

```

Search completed: October 27, 2002, 16:40:50  
 Job time : 198.984 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 15:07:23 : Search time 1487.36 Seconds  
(without alignments)  
10805.408 Million cell updates/sec

Title: US-09-677-374-5

Sequence: 1 atgtcagttgaattctacaa.....agggtgattagcaccgaaaaa 768

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

GenBank: 1: gb\_da: 2: gb\_htg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vi: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_or: 22: em\_ov: 23: em\_pat: 24: em\_ph: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vi: 30: em\_htg\_hum: 31: em\_htg\_inv: 32: em\_htg\_other: 33: em\_htgo\_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	295	38.4	489	6	AX252413	AX252413 Sequence
2	295	38.4	4983	1	AF184152	AF184152 Piscirickettsia
3	270.2	35.2	486	6	AR001082	AR001082 Sequence
4	270.2	35.2	486	6	AR001083	AR001083 Sequence
5	270.2	35.2	486	6	AR027448	AR027448 Sequence
6	270.2	35.2	486	6	AR027449	AR027449 Sequence
7	270.2	35.2	486	6	AR058257	AR058257 Sequence
8	270.2	35.2	486	6	AR058258	AR058258 Sequence
9	270.2	35.2	486	6	118514	118514 Sequence 1
10	270.2	35.2	486	6	118515	118515 Sequence 3
11	270.2	35.2	486	6	166614	166614 Sequence 1
12	270.2	35.2	486	6	166615	166615 Sequence 1
13	270.2	35.2	486	6	188789	188789 Sequence 3
14	270.2	35.2	486	6	188790	188790 Sequence 3
15	270.2	35.2	499	6	AR036204	AR036204 Sequence
16	270.2	35.2	1146	6	AR151733	AR151733 Sequence
17	270.2	35.2	1146	6	AX250579	AX250579 Sequence
18	270.2	35.2	6760	1	CLOCBP	M73817 Clostridium
19	76.4	9.2	6240	1	AB004845	AB004845 Clostridi
20	70.8	9.2	4919	1	CCU40345	U40345 Clostridium
21	70.2	9.1	1069	12	AF283517	AF283517 Synthetic
22	70.2	9.1	1482	6	AR129965	AR129965 Sequence
23	70.2	9.1	2400	1	CTCIPBA	X68233 C.thermoce
24	70.2	9.1	5573	1	CLOCIPA	L08665 Clostridium
25	66.8	8.7	10898	1	AE007606	AE007606 Clostridi
26	61.6	8.0	6555	1	AF155197	AF155197 Acetivibr
27	51	6.6	3048	1	CLORECASE	L04735 Clostridium
28	51	6.6	10782	1	CTH275974	AJ275974 Clostridi
29	47.4	6.2	3360	1	CSCEL2	X55299 C.stercor
30	46.6	6.1	537	1	RIR17KCA	D16515 Rickettsia
31	46	6.0	8246	6	AX345104	AX345104 Sequence
32	45.8	5.9	3120	1	CSCEL72	269359 C.stercor
33	45	5.9	237523	1	RPMX04	AJ235273 Rickettsi
34	45	5.8	1011	1	CTH420770	AJ420770 Clostridi
35	44.4	5.8	10595	6	AX251295	AX251295 Sequence
36	44.4	5.7	3632	6	AX082874	AX082874 Sequence
37	44	5.7	45545	2	AC095545	AC095545 Rattus no
38	44	5.7	125020	9	AF429315	AF429315 Homo sapl
39	44	5.7	186622	2	AC023433	AC023433 Homo sapl
40	43.8	5.7	197164	9	AC064865	AC064865 Homo sapl
41	43	5.6	291050	1	AP000982	AP000982 Sulfolobu
42	43	5.6	72487	9	AL603865	AL603865 Human DNA
43	42.8	5.6	189506	2	AC104102	AC104102 Mus muscu
44	42.8	5.6	203050	1	AL646071	AL646071 Ralstonia
45	42.8	5.6	203050	1	AL646071	AL646071 Ralstonia

#### ALIGNMENTS

RESULT 1  
AX252413  
LOCUS AX252413  
DEFINITION Sequence 5 from Patent WO0168865.  
ACCESSION AX252413  
VERSION AX252413.1 GI:15985721  
KEYWORDS  
SOURCE  
ORGANISM  
Piscirickettsia salmonis.  
Piscirickettsia salmonis  
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;  
Piscirickettsia.  
REFERENCE  
1 (bases 1 to 489)  
AUTHORS Simard,N., Brouwers,H., Jones,S.F., Griffiths,S., Valenzuela,P. and Burzio,L.  
TITLE Fish vaccine against piscirickettsia salmonis  
JOURNAL Patent: WO 0168865-A 5 20-SEP-2001.  
Aqua Health (Europe) Limited (GB)  
FEATURES  
source  
1..489  
/organism="Piscirickettsia salmonis"  
/db\_xref="taxon:1238"  
BASE COUNT 139 a 79 c 144 g 127 t  
ORIGIN

Query Match	38.4%;	Score 295;	DB 6;	Length 489;
Best Local Similarity	76.0%;	Pred. No. 1.5e-64;		
Matches 364;	Conservative 0;	Mismatches 115;	Indels 0;	Gaps 0;

290 GTGTTGCCTCGACGGCCAGCTCTGATCATTAATCTGTGTTTCCCTGGTGGGTGCGCC 349

8 GAGGATGTTTGCAGGTAAGTCTAATTATATCAGTGTGTTTACGTGGCTGTGCCC 67

350 AGACTTCAGCCGCCAGGAAGTTGGCCGCCACCGTGGGTTGTGGCGGTGGCCG 409

68 AGACTT TAGTCGTC AAGAAGTC GAGCTG CCACTGGGCGCTGTTGGCGGTTGCTG 127

[illegible][illegible][illegible]

470 GCGGCTCTGATTGGCTCTAAATCGGTGAGAGCATGACACGACAGGATAAATCAACTGA 529

188 GTGATTATTGGTTCTAAATCGGTCAATCGATGGATCAGCAGGATAAATAAAGCTAA 247

530 ACCAGTCTGTGGAAGAAGTGAAGCCGGCCAGGTTACTTCGTTGGCGTAATCCGGACACCG 589

248 ACCAGAGTTGGAAAAGGTAAGACAGGGCAAGTGACACGTTGGCGTAATCCAGATACAG 307

590 GTACAGCTACTCTGTGGAAACCGGTTCCACCTACCAAGCGTTACAACAACAGGACGCC 649

308 GCATAGTTATAGTCTTGAGCCGCGCTTACCTTACCGCCCTTACAAATACCGACACCCCTC 367

[illegible][illegible]

388 GCGAGCAATATGTGAGAAATTTCAGCAAAAGGCGATGATTCAGGGCAGAGACGAAGAGA 427

Accession	Gene	Species	Accession	Gene	Species
AF184152	4983 bp	DNA	linear	BCT 01-OCT-2000	
AF184152	Piscirickettsia salmonis	alanine racemase (alr) gene, partial cds:			

SOURCE ORGANISM  
Piscirickettsia salmonis.  
Piscirickettsia salmonis  
Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group; Piscirickettsia.

```

/evidence=not_experimental
/product="alanine racemase"
/protein_id="AAG17001.1"
/db_xref="gi:10441347"
/translation="NSAIAINFSEYEDYDVRGSLMIGISPPADKNCIDLPIQGVHC
VSRISLQQLRGSGSYSGATNGCPEDMDQGLISLGYDGGTPIPLAASTPPLNGQRC
ALIGVSNMDIAIDLRKCPAGVGEAVTWGDDLEVEELATRVGIIATELCCNPLPLA

```

```
complement(526. .1257)
/gene="bax"
complement(526. .1257)
```

```

3/codon_start=1
4/transl_table=1
5/evidence=not_experimental
6/product="BAX"
7/protein_id="Agi7002.1"
8/db_xref="GI:10441348"
9/translation="MSLVAITPAVLAQQLPFSAPTSKTOITNFAVSODISINIOEFTI
10LIRKQSEMTINQILAKRQHLISLIDKSSKREITOLAKQYKINLHPKHQDES
11LIRKQSEMTINQILAKRQHLISLIDKSSKREITOLAKQYKINLHPKHQDES
12MEVANNYSMTASRAVAILHTLNTLALQALRDRAHMRNHODVSAFRLAEIGLSTIK
13GKTVLQGLITLIRLSTKSTQIP
1373. 2743

```

```

/evl.dence=experimental
1373. .2743

```

[illegible]

```
DALATL"
2834. .3322
/gene="ospA"
/evidence=experimental
```

```

/note="similar to the rickettsial 17 kDa antigen"
/codon_start=1
/translation_table=11
/product="17 kDa antigen"
/protein_id="AAG17000.1"
/db_xref="GI:10441346"
/translation="NNRGCLQSSLLTIIISFLVCAONFSEQYGAATGAVGVAGQ
LFGSGSVAMATGAGVAVLGLSGIKGSMDDQDITKINQSLFKVAKGQVTRNRNDT
GNSVSEVPVMTYORYNKORRROOYCFEFOKAMLAGOKOEIYGTACRDPGRMOVIST
EK"
3560..4849
/organism="Piscirickettsia salmonis"
/strain="LP-89"
/db_xref="taxon:11238"
/clone="pB12"
/insertion_seq="insertion sequence-like element"
3560..3846
complement(3842..4561)
/gene="tnpA"
complement(3842..4561)

```

```
/gene="tpa"  
/codon_start=1  
/transl_table=1  
/product="transposase"  
/protein_id="AAG16999."  
/db_xref="GI:10441345"
```

VERSION	AR027448.1	GI:5938268
KEYWORDS		
SOURCE	Unknown.	

ORGANISM	Unclassified.	REFERENCE	1 (bases 1 to 486)	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.	Methods of detection using a cellulose binding domain fusion product	JOURNAL	Patent: US 5856201-A 1 05-JAN-1999;	FEATURES	Location/Qualifiers	source	1..486	/organism="unknown"	BASE COUNT	194 a	89 c	79 g	124 t	ORIGIN
Query Match	35.2%;	Score 270.2;	DB 6;	Length 486;	Best Local Similarity	98.9%;	Pred. No. 3.2e-58;	Matches 272;	Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;					
Db	1	ATGTCAGTTGAATCTCTACACCTCTAACCAATCAGACAAACAACTCAATTACACCAATA	60		16	ATGTCAGTTGAATTTTACAACTCTAACCAATCAGACAAACAACTCAATTACACCAATA	75											
Qy	61	ATCAAAATTACTAACACATCTGCACAGTATTAAATTTAAATGACGTAAAGTTAGATAT	120		76	ATCAAAATTACTAACACATCTGCACAGTATTAAATTTAAATGACGTAAAGTTAGATAT	135											
Db	121	TATTTACACAGTATGATGATACACACAGGACAACTTCTGCTGTGACCAATGCTGGTGCATTA	180		136	TATTTACACAGTATGATGATACACACAGGACAACTTCTGCTGTGACCAATGCTGGTGCATTA	195											
Qy	181	TTAGCAAAATGATGTTGATGATTAACACTGTGCAAGAGACAGCAAACTTGTGTTAAAGAAACA	240		196	TTAGCAAAATGATGTTGATGATTAACACTGTGCAAGAGACAGCAAACTTGTGTTAAAGAAACA	255											
Qy	241	GCAAGCCCAACATCAACCTATGATACATATCTGCA	275		256	GCAAGCCCAACATCAACCTATGATACATATGTTGA	290											
RESULT 6	AR027449/c	486 bp	DNA	Linear	PAT 29-SEP-1999													
LOCUS	AR027449	486 bp	DNA	Linear	PAT 29-SEP-1999													
DEFINITION	Sequence 3 from patent US 5856201.																	
ACCESSION	AR027449																	
VERSION	AR027449.1	GI:5938269																
KEYWORDS	Unknown.																	
SOURCE	Unknown.																	
ORGANISM	Unclassified.																	
REFERENCE	1 (bases 1 to 486)																	
AUTHORS	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.																	
TITLE	Methods of detection using a cellulose binding domain fusion product																	
JOURNAL	Patent: US 5856201-A 3 05-JAN-1999;																	
FEATURES	Location/Qualifiers																	
source	1..486																	
BASE COUNT	124 a	79 c	89 g	194 t														
ORIGIN	1	ATGTCAGTTGAATCTCTACACCTCTAACCAATCAGACAAACAACTCAATTACACCAATA	60		16	ATGTCAGTTGAATTTTACAACTCTAACCAATCAGACAAACAACTCAATTACACCAATA	75											
Query Match	35.2%;	Score 270.2;	DB 6;	Length 486;	Best Local Similarity	98.9%;	Pred. No. 3.2e-58;	Matches 272;	Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;					
Db	471	ATGTCAGTTGAATTTTACAACTCTAACCAATCAGACAAACAACTCAATTACACCAATA	412		61	ATCAAAATTACTAACACATCTGCACAGTATTAAATTTAAATGACGTAAAGTTAGATAT	120											
Qy	411	ATCAAAATTACTAACACATCTGCACAGTATTAAATTTAAATGACGTAAAGTTAGATAT	352		121	TATTTACACAGTATGATGATACACACAGGACAACTTCTGCTGTGACCAATGCTGGTGCATTA	180											

Db	351	TATTACACAAGTGAATGTCATACCAAGGACAACAACTTTCTGTGTGGACCATCGTGTCATTA	292
OY	181	TTAGAATAAGCTATGTTGATTAACACTAAGCAAAAGTGACAGCAAACCTTGGTTAAGAACCA	240
Db	291	TTAGAAAATAGCTATGTTGATTAACACTAAGCAAAAGTGACAGCAAACCTTGGTTAAGAACCA	232
OY	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
Db	231	GCAAGCCCAACATCAACCTATGATACATATCTTGA	197
RESULT 7			
LOCUS	AR058257	486 bp	DNA
DEFINITION	Sequence 1 from patent US 5837814.		linear
ACCESSION	AR058257		
VERSION	AR058257.1	GI:5983834	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.		
TITLE	Cellulose binding domain proteins		
JOURNAL	Patent: US 5837814-A 1 17-NOV-1998;		
FEATURES	Location/Qualifiers		
source	1..486		
BASE COUNT	194 a	89 c	79 g
ORIGIN	/organism="unknown"		124 t
Query Match	35.2%;	Score 270.2;	DB 6; Length 486;
Best Local Similarity	98.9%;	Pred. No. 3.2e-58;	
Matches 272;	Conservative 0;	Mismatches 3;	Indels 0; Gaps 0;
OY	1	ATGCAGTGTGAATCTCTACACTCTAACAAATCAGACAACAACTCAATTACACCAATA	60
Db	16	ATGTAGTGTGAATTTTACAACTCTACAAATCAGACAACAACTCAATTACACCAATA	75
OY	61	ATCAAAATTACTTACACATCTGCACAGTAGTTAAATTTAAATGAGCTAAAAAGTTAGATAT	120
Db	76	ATCAAAATTACTTACACATCTGCACAGTAGTTAAATTTAAATGAGCTAAAAAGTTAGATAT	135
OY	121	TATTACACAAGTGAATGTCATACCAAGGACAACCTTCTGTGTGGACCATCGTGTGCATTA	180
Db	136	TATTACACAAGTGAATGTCATACCAAGGACAACCTTCTGTGTGACCATCGTGTGCATTA	195
OY	181	TTAGAATAAGCTATGTTGATTAACACTAAGCAAAAGTGACAGCAAACCTTGGTTAAGAACCA	240
Db	196	TTAGAATAAGCTATGTTGATTAACACTAAGCAAAAGTGACAGCAAACCTTGGTTAAGAACCA	255
OY	241	GCAAGCCCAACATCAACCTATGATACATATCTGGA	275
Db	256	GCAAGCCCAACATCAACCTATGATACATATCTTGA	290
RESULT 8			
LOCUS	AR058258/c	486 bp	DNA
DEFINITION	Sequence 3 from patent US 5837814.		linear
ACCESSION	AR058258		
VERSION	AR058258.1	GI:5983835	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.		
TITLE	Cellulose binding domain proteins		
JOURNAL	Patent: US 5837814-A 3 17-NOV-1998;		
FEATURES	Location/Qualifiers		
source	1..486		
	/organism="unknown"		



BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN  
Query Match 35.2%; Score 270.2; DB 6; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.2e-58;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 60  
DB 471 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 412  
QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTAGATAT 120  
DB 411 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTAGATAT 352  
QY 121 TATTACACAAAGTGATGTACACAAAGCAAACTTCTGCTGTGACCATGCTGGTGCAATTA 180  
DB 351 TATTACACAAAGTGATGTACACAAAGCAAACTTCTGCTGTGACCATGCTGGTGCAATTA 292  
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240  
DB 291 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 232  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275  
DB 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197  
RESULT 9  
118514 LOCUS 486 bp DNA linear PAT 07-OCT-1996  
DEFINITION Sequence 1 from patent US 5496934.  
ACCESSION 118514  
VERSION 118514.1 GI:1598869  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shiplegl,I., Goldstein,M.A. and Doi,R.H.  
TITLE Nucleic acids encoding a cellulose binding domain  
JOURNAL Patent: US 5496934-A 1 05-MAR-1996;  
FEATURES  
source 1..486  
BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN  
Query Match 35.2%; Score 270.2; DB 6; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.2e-58;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 60  
DB 16 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 75  
QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTAGATAT 120  
DB 76 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTAGATAT 135  
QY 121 TATTACACAAAGTGATGTACACAAAGCAAACTTCTGCTGTGACCATGCTGGTGCAATTA 180  
DB 136 TATTACACAAAGTGATGTACACAAAGCAAACTTCTGCTGTGACCATGCTGGTGCAATTA 195  
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240  
DB 196 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 255  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275  
DB 256 GCAAGCCCAACATCAACCTATGATACATATGTTGA 290

RESULT 10  
118515/c LOCUS 486 bp DNA linear PAT 07-OCT-1996  
DEFINITION Sequence 3 from patent US 5496934.  
ACCESSION 118515  
VERSION 118515.1 GI:1598870  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shiplegl,I., Goldstein,M.A. and Doi,R.H.  
TITLE Nucleic acids encoding a cellulose binding domain  
JOURNAL Patent: US 5496934-A 3 05-MAR-1996;  
FEATURES  
source 1..486  
BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN  
Query Match 35.2%; Score 270.2; DB 6; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.2e-58;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 60  
DB 471 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 412  
QY 61 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTAGATAT 120  
DB 411 ATCAAAATTTACTACACATCTGACAGTGATTTAAATTAATGACGTAAGTAGATAT 352  
QY 121 TATTACACAAAGTGATGTACACAAAGCAAACTTCTGCTGTGACCATGCTGGTGCAATTA 180  
DB 351 TATTACACAAAGTGATGTACACAAAGCAAACTTCTGCTGTGACCATGCTGGTGCAATTA 292  
QY 181 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 240  
DB 291 TTAGGAATAGCTATGTTGATACACTAGCAAAAGTGACAGCAAACTTCGTTAAAGAAACA 232  
QY 241 GCAAGCCCAACATCAACCTATGATACATATCTGGA 275  
DB 231 GCAAGCCCAACATCAACCTATGATACATATGTTGA 197  
RESULT 11  
166614 LOCUS 486 bp DNA linear PAT 29-DEC-1997  
DEFINITION Sequence 1 from patent US 5670623.  
ACCESSION 166614  
VERSION 166614.1 GI:2724592  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shiplegl,I., Goldstein,M.A. and Doi,R.H.  
TITLE Methods of use of cellulose binding domain proteins  
JOURNAL Patent: US 5670623-A 1 23-SEP-1997;  
FEATURES  
source 1..486  
BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN  
Query Match 35.2%; Score 270.2; DB 6; Length 486;  
Best Local Similarity 98.9%; Pred. No. 3.2e-58;  
Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 60  
DB 16 ATGTCAGTTGAATTTCTACAACTCTAACAAATCAGCACAACAACTCAATTAACACATA 75

Qy	61	ATCAAAATTTACTACACACATCTGCACGTATTTAAATTTAAATGACGTAAAGTTGATAT	120
Db	76	ATCAAAATTTACTACACACATCTGCACGTATTTAAATTTAAATGACGTAAAGTTAGATAT	135
Qy	121	TATTAACACAAGTGATGTGTACACACAAGACAACCTTCTGTGTGACACAGTCGTGTCATTA	180
Db	136	TATTAACACAAGTGATGTGTACACACAAGACAACCTTCTGTGTGACACAGTCGTGTCATTA	195
Qy	181	TTTAGAANAATAGCTAATGTTGATATACACTAGCAAAAGTGACAGCAAACTTGCTTTAAGAACCA	240
Db	196	TTAGGAANAATAGCTAATGTTGATATACACTAGCAAAAGTGACAGCAAACTTGCTTTAAGAACCA	255
Qy	241	GCAGGCCCAACATCAACCTATGATACATATCTGGA	275
Db	256	GCAGGCCCAACATCAACCTATGATACATATCTGGA	290

RESULT	12				
LOCUS	166615/c				
DEFINITION	I66615	486 bp	DNA	linear	PAT 29-DEC-1997
ACCESSION	Sequence 3 from patent US 5670623.				
VERSION	I66615				
KEYWORDS	I66615.1 GI:2724593				
SOURCE	.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 486)				
TITLE	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.				
JOURNAL	Methods of use of cellulose binding domain proteins				
FEATURES	Patent: US 5670623-A 3 23-SEP-1997;				
	Location/Qualifiers				
source	1..486				
	/organism="unknown"				
BASE COUNT	124 a	79 c	89 g	194 t	
ORIGIN					

Query Match	35.2%	Score 270.2	DB 6	Length 486
Best Local Similarity	98.9%	Pred. No. 3.2e-58		
Matches 272; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY	1	ANGAGTGTGATTTCTACCAAGCTCTAACAAATGAGACAAACAAACTCAATTTCACCAATA	60
Db	471	ANGCTGTTGAATTTTACCACTCTTAACCAATGACGACAAACAACTCAATTTCACCAATA	412
QY	61	ATCAAAATTTACTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTGTGATAT	120
Db	411	ATCAAAATTTACTAACACATCTGACAGTGAATTTAAATTTAAATGAGCTAAAGTGTGATAT	352
QY	121	TATTACACAAAGTATGAGTACACAAAGCAAACTTCTGGTGTGACATCTGGTGATTA	180
Db	351	TATTACACAAAGTATGAGTACACAAAGCAAACTTCTGGTGTGACATCTGGTGATTA	292
QY	181	TTAGGAAATAGCTATGTTTATTAACACTATACCAAGTGACAGCAAACTTCGTTAAAGAAACA	240
Db	291	TTAGGAAATAGCTATGTTTATTAACACTATACCAAGTGACAGCAAACTTCGTTAAAGAAACA	232
QY	241	GCAAGCCCAACATCAACTGATGATCATATTCGGA	275
Db	231	GCAAGCCCAACATCAACTGATGATCATATTCGGA	197

RESULT 13			
LOCUS	188789	486 bp	DNA
DEFINITION	Sequence 1 from patent US 5715044.		linear
ACCESSION	188789		PAT 10-AUG-1996
VERSION	188789.1		
KEYWORDS	GI:3408729		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
	1 (bases 1 to 486)		

AUTHORS	Shoseyov, O., Shpiegel, I., Goldstein, M.A. and Dol, R.H.
TITLE	Cellulose binding domain fusion proteins
JOURNAL	Patent: US 5719044-A 1 17-FEB-1998;
FEATURES	location/Qualifiers
source	1. .486
	/organism="unknown"
BASE COUNT	194 a 89 c 79 g 124 t
ORIGIN	

Query Match	35.2%	Score	270.2	DB	6	Length	486
Best Local Similarly	98.9%	Pred. No.	3.2e-58				
Matches 272; Conservative	0	Mismatches	3	Indels	0	Gaps	0

QY	1	ATTCAGTTGSAATTCCTACAACTCTACAAAAATCAGCAAAACAAACTCAATTACACCATTA	60
QY	16	ATGTAGTTGTAATTTTACAACTCTACAAATAGCACAACAAACTCAATTACACCATTA	75
Db	61	ATCAAAATTTACTAACACATCTGACAGTATTTAAATTTAAATGACGTAAAGTTAGATAT	120
QY	76	ATCAAAATTTACTAACACATCTGACAGTATTTAAATTTAAATGACGTAAAGTTAGATAT	135
Db	121	TATTTACACAAGTATGATGACACAAAGGACAAACTCTCTGCTGACCATGCGTGACATTA	180
QY	136	TATTTACACAAGTATGATGACACAAAGGACAAACTCTCTGCTGACCATGCGTGACATTA	195
Db	181	TTAGGAATTTAGCTATGTTGATATACACTAGCAAAAGTGACAGCAAACTTGTTAAAGAAACA	240
QY	196	TTAGGAATTTAGCTATGTTGATATACACTAGCAAAAGTGACAGCAAACTTGTTAAAGAAACA	255
Db	241	GCAAGCCCAACATCAACTATGATATCATATCTCGGA	275
QY	256	GCAAGCCCAACATCAACTATGATATCATATCTCGGA	290

RESULT 14				
188790/c				
LOCUS	188790	486 bp	DNA	linear
DEFINITION	Sequence 3 from patent US 5719044.			
ACCESSION	188790			PAT 10-AUG-1998
VERSION	188790.1	GI:3408730		
KEYWORDS	.			
SOURCE	unknown.			
ORGANISM	unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 486)			
TITLE	Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Doi,R.H.			
JOURNAL	Cellulose binding domain fusion proteins			
FEATURES	Patent: US 5719044-A 3 17-FEB-1998;			
source	location/Qualifiers			
	1..486			
	/organism="unknown"			
BASE COUNT	124 a	79 c	89 g	194 t
ORIGIN				

Query Match	Similarity	35.2%	Score 270.2	DB 6	Length 486
Best Local	Similarity	98.9%	Pred. No. 3.2e-58		
Matches	272	Conservative	0	Mismatches	3
				Indels	0
				Gaps	0
QY	1	ATGTCAGTTGAATTCCTACACCTCTAACAAANTCAGACACAACAACTCAATTACACCAATA	60		
DB	471	ATGTCAGTTGAATTCCTACACCTCTAACAAANTCAGACACAACAACTCAATTACACCAATA	412		
QY	61	ATCAAAATTTACATACACACATCTGCAGTGTATTAATTTAAATGAGCTTAAATTTGATAT	120		
DB	411	ATCAAAATTTACATACACACATCTGCAGTGTATTAATTTAAATGAGCTTAAATTTGATAT	352		
QY	121	TATTAACACAAGTATGTTGACACAGAAGACAACCTTCTGTGTTGACATGCTGTGTCATTA	180		
DB	351	TATTAACACAAGTATGTTGACACAGAAGACAACCTTCTGTGTTGACATGCTGTGTCATTA	292		
QY	181	TTTAGCAATAGCTATGTTGTATATACACTATGACAAAGTGACAGCAAACTTGTGTTAAAGACA	240		
DB	291	TTTAGCAATAGCTATGTTGTATATACACTATGACAAAGTGACAGCAAACTTGTGTTAAAGACA	232		

QY 241 GCAAGCCCAACATCACTATGATACATATCTGGA 275  
|||||  
Db 231 GCAAGCCCAACATCACTATGATACATATGTTGA 197

## RESULT 15

AR096204 499 bp DNA linear PAT 08-SEP-2000

LOCUS AR096204 Sequence 5 from patent US 6005092.

ACCESSION AR096204

VERSION AR096204.1 GI:10024795

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 499)

AUTHORS Shoseyov, O. and Shaul, Z.

TITLE Arabidopsis thaliana endo-1,4'-beta.-glucanase gene and promoter

JOURNAL Patent: US 6005092-A 5 21-DEC-1999;

FEATURES Location/Qualifiers

source 1..499

BASE COUNT 197 a 93 c 82 g 127 t

ORIGIN

## Query Match

Best Local Similarity 98.9%; Score 270.2; DB 6; Length 499;

Matches 272; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAAGTGAATTTTACAACTCTACAAATCAGACAAACAACTCAATTACCCATTA 60  
|||||  
Db 21 ATGTCAAGTGAATTTTACAACTCTACAAATCAGACAAACAACTCAATTACCCATTA 80  
QY 61 ATCAAAATTACTAACAACATCTGACAGTGATTTAAATTAAATGACGTAAAGTTAGATAT 120  
|||||  
Db 81 ATCAAAATTACTAACAACATCTGACAGTGATTTAAATTAAATGACGTAAAGTTAGATAT 140  
QY 121 TATTACACAAGTGATGTGACACAAGACAACCTTCTGTGTGACCATGCTGGTGATTA 180  
|||||  
Db 141 TATTACACAAGTGATGTGACACAAGACAACCTTCTGTGTGACCATGCTGGTGATTA 200  
QY 181 TTAGAAATAGCTATGTGTGTAACACTAGCAAGTAGACAAACTTCGTTAAGAAACA 240  
|||||  
Db 201 TTAGAAATAGCTATGTGTGTAACACTAGCAAGTAGACAAACTTCGTTAAGAAACA 260  
QY 241 GCAAGCCCAACATCACTATGATACATATCTGGA 275  
|||||  
Db 261 GCAAGCCCAACATCACTATGATACATATGTTGA 295

Search completed: October 27, 2002, 17:46:49  
Job time : 1495.36 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:15:49 : Search time 1538.65 Seconds  
(without alignments)  
6736.842 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atcgtcagtgatcgaattctacaa.....agggtgattgacacgcgaataaa 768

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlmu:\*  
5: em\_estlov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	2.9	709	9	AV760726
2	21	2.7	391	12	AQ252889
3	20	2.6	516	12	AQ518147
4	20	2.6	717	10	BE532389
5	20	2.6	823	10	BI186311
6	19	2.5	184	10	BE247681
7	19	2.5	190	12	FR0011976
8	19	2.5	191	9	AM661978
9	19	2.5	210	10	BF803712
10	19	2.5	237	9	AA909847
11	19	2.5	265	10	244341
12	19	2.5	283	9	AI817750
13	19	2.5	322	9	AI182462
14	19	2.5	335	10	D51829
15	19	2.5	336	10	BE246888
16	19	2.5	348	10	T51729
17	19	2.5	363	9	AA825376

18	19	2.5	384	10	BF109466
19	19	2.5	456	9	AI559589
20	19	2.5	459	9	AV763598
21	19	2.5	475	9	AM068733
22	19	2.5	502	10	BM309775
23	19	2.5	508	9	AI478023
24	19	2.5	510	9	AI678459
25	19	2.5	527	10	BM381932
26	19	2.5	528	9	AM658421
27	19	2.5	536	9	AI637152
28	19	2.5	630	10	BM335231
29	19	2.5	645	10	BC924401
30	19	2.5	648	12	BH381604
31	19	2.5	652	9	AI267409
32	19	2.5	670	10	BI151293
33	19	2.5	678	9	AM167360
34	19	2.5	697	10	BM350333
35	19	2.5	719	10	BI868632
36	19	2.5	744	10	BI764273
37	19	2.5	816	12	BH524872
38	19	2.5	846	10	BI667298
39	19	2.5	892	9	AU135371
40	19	2.5	919	12	CNS04Q0J
41	19	2.5	2405	10	BC715056
42	18	2.3	291	9	AV214643
43	18	2.3	342	12	CNS005SM
44	18	2.3	372	12	BH395820
45	18	2.3	382	9	AM588379

#### ALIGNMENTS

RESULT 1	AV760726	709 bp	MRNA	linear	EST 19-OCT-2000
LOCUS	AV760726	MDS	Homo sapiens	CDNA clone	MDSAJE03 5', mRNA sequence.
DEFINITION	AV760726				
ACCESSION	AV760726.1	GI:10918574			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 709)  
Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,  
,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G.,  
Yang,Y., Gao,G., Zhang,O., Chen,S., Han,Z. and Chen,Z.  
Homo sapiens CDNA MDS clones

Unpublished (2000)  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919 (ex.45)  
Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.  
Location/Qualifiers

1. 709

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="MDSAJE03"

/clone\_lib="MDS"

/tissue\_type="Bone marrow"

/cell\_type="CD34+ hematopoietic stem/progenitor cell"

/lab\_host="BM25.8"

/note="Vector: pTriplet2; Site\_1: sf11a; Site\_2: sf11b"

BASE COUNT 172 a 159 c 188 g 190 t  
Query Match 2.9%; Score 22; DB 9; Length 709;

Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 147 ACAACTTCTGTGTCACCAT 168  
|||||  
Db 497 ACAACTTCTGTGTCACCAT 518

RESULT 2  
AQ252889/c 391 bp DNA linear GSS 10-OCT-1998  
LOCUS HS\_3029\_A2\_D03\_MR CIT Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone Plate=3029 Col=6 Row=G, DNA sequence.  
ACCESSION AQ252889  
VERSION AQ252889.1 GI:3723803  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 391)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3029 row: G column: 6  
Class: BAC ends  
High quality sequence stop: 391.  
Location/Qualifiers  
1..391  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3029 Col=6 Row=G"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 77 a 73 c 85 g 156 t

ORIGIN

Query Match 2.7%; Score 21; DB 12; Length 391;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 AATCAGCACAACAACCTCAA 49  
|||||  
Db 153 AATCAGCACAACAACCTCAA 133

RESULT 3  
AQ518147 516 bp DNA linear GSS 05-MAY-1999  
LOCUS HS\_5090\_A1\_F08\_T7A RPT-11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plate=666 Col=15 Row=K, DNA sequence.  
ACCESSION AQ518147  
VERSION AQ518147.1 GI:4743329  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 516)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPT-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Research Genetics (Info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 666 row: K column: 15  
Seq primer: 17  
Class: BAC ends  
High quality sequence stop: 516.  
Location/Qualifiers  
1..516  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=666 Col=15 Row=K"  
/clone\_lib="RPT-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"

BASE COUNT 122 a 156 c 129 g 104 t 5 others

ORIGIN

Query Match 2.6%; Score 20; DB 12; Length 516;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 296 GCTCGCAGGCGACCTCTCTG 315  
|||||  
Db 191 GCTCGCAGGCGACCTCTCTG 210

RESULT 4  
BE532389 717 bp mRNA linear EST 09-AUG-2000  
LOCUS 601233730F1 NCL\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3597462 5'  
DEFINITION mRNA sequence.  
ACCESSION BE532389  
VERSION BE532389.1 GI:9761034  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 717)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
http://image.llnl.gov

Plate: LAM8776 row: 1 column: 07  
High quality sequence stop: 510.

# FEATURES

Location/Qualifiers

1. 717

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3597462"

/clone\_1lb="NCI CGAP Mam6"

/sex="female, virgin"

/tissue\_type="infiltrating ductal carcinoma"

/dev\_stage="5 months"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI. Cloned unidirectionally. Primer: oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 228 a 152 c 150 g 187 t

# ORIGIN

Query Match 2.6%; Score 20; DB 10; Length 717;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 ACAATCAGCACAACAAC 45  
Db 462 ACAATCAGCACAACAAC 481

# RESULT 5

BI186311/c 823 bp mRNA linear EST 10-JUL-2001

DEFINITION UML-P-FN-cf-b-07-0-UML.52 UML-P-FN Sus scrofa cDNA clone

UML-P-FN-cf-b-07-0-UML.3, mRNA sequence.

BI186311

BI186311.1 GI:14660720

EST.

KEYWORDS

EST.

ORGANISM

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

Caetano, A.R., Johnson, R.K. and Pomp, D.

Generation and sequence characterization of a normalized cDNA

library from swine ovarian follicles

unpublished (2001)

Contact: Pomp, D

Department of Animal Science

University of Nebraska, Lincoln

Lincoln, NE 68583-0908, USA

Tel: 402 472 6416

Fax: 402 472 6362

Email: dpomp@unl.edu

Oligo-dt track not found, Not I site shown in beginning of sequence

is likely internal to the message.

Seq primer: M13 -29

POLYA-No.

Location/Qualifiers

1. 823

/organism="Sus scrofa"

/strain="University of Nebraska, Lincoln Swine Selection

lines"

/db\_xref="taxon:9823"

/clone="UML-P-FN-cl-b-07-0-UML"

/clone\_1lb="UML-P-FN"

/dev\_stage="ADULT"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; Site:1: Not I; Site:2: Eco RI; The UML-P-FN

library is a normalized library representing porcine

ovarian follicles, ranging between 2.0 to 10.0 mm in

diameter, collected during 7 days of the follicular phase

of the pig estrous cycle. This library was derived from

the library UML-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

PAC\_SEQ=None found"

BASE COUNT 151 a 272 c 231 g 167 t 2 others

# ORIGIN

Query Match 2.6%; Score 20; DB 10; Length 823;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 293 GTTCGTCGACGACGCTCT 312  
Db 730 GTTCGTCGACGACGCTCT 711

# RESULT 6

BE247681/c 184 bp mRNA linear EST 03-OCT-2001

DEFINITION TCBAPE25356 Pediatric pre-B cell acute lymphoblastic leukemia

Baylor-HGSC project-TCBA Homo sapiens cDNA clone TCBAPE25356, mRNA

sequence.

BE247681

BE247681.1 GI:9099528

EST.

KEYWORDS

EST.

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman Jr., F.R., Muzny, D.

, Bouck, J., Gibbs, R.A. and Margolin, J.F.

Pediatric Leukemia cDNA Sequencing Project

Unpublished (2000)

Contact: Dr. Judith F. Margolin

Texas Children's Cancer Center and Human Genome Sequencing Center

at Baylor College of Medicine

1102 Bates, MC3-3320 Houston, TX 77030, USA

Tel: 832-824-4536

Fax: 832-825-4038

Email: clones@ccc.org

Citation: Garinchi, P. and Hayashizaki, Y. High efficiency

full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Seq primer: M13 primer.

Location/Qualifiers

1. 184

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TCBAPE25356"

/clone\_1lb="Pediatric pre-B cell acute lymphoblastic

leukemia Baylor-HGSC project-TCBA"

/sex="male"

/tissue\_type="leukopheresis"

/cell\_type="pre-B cell"

/dev\_stage="pediatric 2 years"

/lab\_host="DH10B"

/note="Vector: lambda PSB; Site:1: BamHI; Site:2: EcoRI;

First strand cDNA was primed with an anchored

XhoI-oligo(dT) primer [5'GGAGAGTCGAGCGCGGAGAG(T)VN

3'; V-A, C, G, N-A, C, G, T] and then dg tailed. Second strand

was primed with a BamHI-dc primer

[5'AGAGAGTCGAGTCGAGCGCGGCAATTAATATAT(C) 3'].

Double-stranded cDNA was then digested with BamHI and XhoI

and directionally cloned into the BamHI and SalI sites of

lambda psb vector. Library went through one round of

normalization. Library was constructed by Wei Yu at RIKEN

of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,

Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,

Schneider C, Hayashizaki Y, High efficiency selection of

full-length cDNA by improved biotinylated cap trapper.,

DNA Res 4: 1, 61-66, Feb 28, 1997]"

```
BASE COUNT      43 a      27 c      30 g      79 t      5 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 19; DB 10; Length 184;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ACAGTGATTAAATTAA 101
Db 31 ACAGTGATTAAATTAA 13

RESULT 7
LOCUS FR0011976 190 bp DNA linear GSS 18-SEP-1997
DEFINITION F.rubripes GSS sequence, clone 072L05AE5, genomic survey sequence.
ACCESSION AF003232.1 GI:2448802
VERSION AF003232.1 GI:2448802
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 190)
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y.,
Williams,G. and Brenner,S.
Direct Submission
Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biolhelp@hmp.mrc.ac.uk
Vector: m13mp18
V-type: phage
PRIMER: M13
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES
source location/Qualifiers
1..190
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 072L05"
/clone="072L05AE5"
BASE COUNT      56 a      38 c      40 g      56 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 19; DB 12; Length 190;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 CAGGCAGCTCTGATCA 319
Db 119 CAGGCAGCTCTGATCA 137

RESULT 8
LOCUS AW661978 191 bp mRNA linear EST 06-APR-2000
DEFINITION h117h11.x1 NCI-CGAP-GUI Homo sapiens cDNA clone IMAGE:2972613 3'
similar to contains element TAR1 repetitive element ;, mRNA
sequence.
ACCESSION AW661978
VERSION AW661978.1 GI:7454514
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 191)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/lresources.shtml
Seq primer: -400p from Glbco
High quality sequence stop: 157.
FEATURES
source location/Qualifiers
1..191
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2972613"
/clone_lib="NCI-CGAP-GUI"
/tissue_type="2 pooled high-grade transitional cell
tumors"
/lab_host="DH10B"
/note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
Site.1: SalI; Site.2: NotI; Cloned unidirectionally.
Primer: oliigo dt. Library constructed by Life
Technologies."
BASE COUNT      70 a      32 c      35 g      54 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 19; DB 9; Length 191;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ACAGTGATTAAATTAA 101
Db 32 ACAGTGATTAAATTAA 50

RESULT 9
LOCUS BF803712 210 bp mRNA linear EST 12-JAN-2001
DEFINITION CM0-C10136-021100-670-d11 C10136 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF803712
VERSION BF803712.1 GI:12132701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 210)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
Nagai,M.A., de Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&cl=CM0-C10136-
021100-670-d11&ts=2000-11-02&td=1)
Seq primer: puc 18 forward
High quality sequence start: 6
```



High quality sequence stop: 210.  
Location/Qualifiers  
1. 210  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="C10136"  
/dev\_stage="Adult"  
/note="Organ: colon;ins: Vector: puc18; Site:1: SmaI;  
Site:2: SmaI. A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 95 a 35 c 24 g 55 t 1 others

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 210;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 83 ACAGTGATTTAAATTAA 101  
|||||  
Db 174 ACAGTGATTTAAATTAA 192

RESULT 10  
AA909847 237 bp mRNA linear EST 26-AUG-1998  
LOCUS O101910.s1 NCI\_CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1522242 3',  
DEFINITION mRNA sequence.  
ACCESSION AA909847  
VERSION AA909847.1 GI:3047549  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 237)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/db/rp/image/image.html  
Insert Length: 528 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 210.  
Location/Qualifiers  
1. 237  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1522242"  
/clone\_1ib="NCI\_CGAP\_Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker. 1st strand cDNA was prepared from  
neuroendocrine lung carcinoid, and was then primed with a  
Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library is normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 91 a 41 c 42 g 63 t

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 237;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 83 ACAGTGATTTAAATTAA 101  
|||||  
Db 31 ACAGTGATTTAAATTAA 49

RESULT 11  
244341 265 bp mRNA linear EST 14-NOV-1994  
LOCUS HSC1YD071 normalized infant brain cDNA Homo sapiens cDNA clone  
DEFINITION c-1y007, mRNA sequence.  
ACCESSION 244341  
VERSION 244341.1 GI:573469  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 265)  
Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes  
Mitchell,H., Marjage-Samson,R., Pictu,G., Pouliot,Y.,  
Sebastiani,Kabakchis,C. and Tessier,A.  
IMAGE: molecular integration of the analysis of the human genome  
and its expression  
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
95277534  
JOURNAL  
MEDLINE  
COMMENT Contact: Genethon  
Genethon Centre de recherche sur le Genome Humain  
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
Tel: 33169472800  
Fax: 33160778698  
Email: genexpress@genethon.fr  
Single read.  
Genexpress\_library\_idt: C; Genexpress\_sequence\_idt: ylc-lyd07  
Seq primer: (-21)M13 universal.  
Location/Qualifiers  
1. 265  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="c-1y07"  
/clone\_1ib="normalized infant brain cDNA"  
/sex="Female"  
/tissue\_type="total brain"  
/dev\_stage="3 months old"  
/note="Organ: brain; Vector: lafmid BA; Site:1: HindIII;  
Site:2: NotI; sex=Female; dev\_stage=3 months old;  
isolate=muscular atrophy patient; tissue\_type=total brain  
; total mRNA was oligo-(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lafmid BA vector. Clone library from B. Soares, Psychiatry  
Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S in press"

BASE COUNT 76 a 43 c 48 g 89 t 9 others

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 265;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 83 ACAGTGATTTAAATTAA 101  
|||||  
Db 209 ACAGTGATTTAAATTAA 191

RESULT 12  
A1817750

LOCUS A1817750 283 bp mRNA linear EST 21-DEC-1999  
 DEFINITION wk25909.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2413408 3',  
 mRNA sequence.  
 ACCESSION A1817750  
 VERSION A1817750.1 GI:5436829  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 283)  
 NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute / National Institute of Neurological  
 Disorders and Stroke, Brain Tumor Genome Anatomy Project  
 (CGAP/BRGAP), Tumor Gene Index  
 Unpublished (1998)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbrrp/image/image.html](http://www-bio.llnl.gov/dbrrp/image/image.html)  
 Insert length: 346 Std Error: 0.00  
 Seq primer: -40UP from Gibco.  
 Location/Qualifiers  
 1..283  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2413408"  
 /clone\_1bp="NCI CGAP\_Brn25"  
 /tissue\_type="anaplastic oligodendroglioma"  
 /lab\_host="DH10B"  
 /note="Organ: Brain; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer (5'  
 TGTTCACATCTGACGTGAGCGGCGCCGACATGTTTTTTTTTTTTTTTT  
 T 3'); double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."

BASE COUNT 96 a 62 c 46 g 79 t  
 ORIGIN  
 Query Match 2.5%; Score 19; DB 9; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 ACAGTATTAAATTAA 101  
 |||||||  
 Db 36 ACAGTATTAAATTAA 54

RESULT 13  
 A1182462/c 322 bp mRNA linear EST 08-OCT-1998  
 LOCUS A1182462  
 DEFINITION uc26h10.r1 Soares\_mammary\_gland\_NbMNG Mus musculus cDNA clone  
 IMAGE:1399171 5', mRNA sequence.  
 ACCESSION A1182462  
 VERSION A1182462.1 GI:3733100  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 322)  
 Mammaia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

TITLE Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 The Washu-HMMI Mouse EST Project  
 Unpublished (1996)  
 JOURNAL Contact: Marra M/Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:910887  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 303.  
 Location/Qualifiers  
 1..322  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1399171"  
 /clone\_1bp="Soares\_mammary\_gland\_NbMNG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia  
 ) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer (5'  
 TGTTCACATCTGACGTGAGCGGCGCCGACATGTTTTTTTTTTTTTTTT  
 T 3'); double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M.Fatima  
 Bonaldo."

BASE COUNT 47 a 100 c 95 g 79 t 1 others  
 ORIGIN  
 Query Match 2.5%; Score 19; DB 9; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 TGGTACACAGCAAACT 153  
 |||||||  
 Db 185 TGGTACACAGCAAACT 167

RESULT 14  
 D51829/c 335 bp mRNA linear EST 14-SEP-1995  
 LOCUS D51829  
 DEFINITION HUM053G02B Clontech human fetal brain polyA+ mRNA (#535) Homo  
 sapiens cDNA clone GEN-053G02 5', mRNA sequence.  
 ACCESSION D51829  
 VERSION D51829.1 GI:952065  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 335)  
 Fujikawa, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,  
 Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi,  
 A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H.,  
 Shin, S. and Nakamura, Y.  
 Fujikawa et al. (1995)  
 Unpublished (1995)  
 Contact: Tsutomu Fujikawa  
 Otsuka GEN Research Institute  
 Otsuka Pharmaceutical Co., Ltd

463-10 Kagasuno Kawachi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035

## FEATURES

## source

Location/Qualifiers

1. .335

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="GEN-053602"

/clone\_lib="Clontech human fetal brain polyA+ mRNA (#6535

)"

BASE COUNT

93 a 44 c 58 g 126 t 14 others

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 335;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ACAGTGATTAAATTAA 101

Db 95 ACAGTGATTAAATTAA 77

RESULT 15

BE246888 336 bp mRNA linear EST 03-OCT-2001

LOCUS TCBAP15756 Pediatric pre-B cell acute lymphoblastic leukemia

DEFINITION Baylor-HGSC project-TCBA Homo sapiens CDNA clone TCBAP5756; mRNA

SEQUENCE

BE246888

BE246888.1 GI:9098639

KEYWORDS

SOURCE

ORGANISM

human

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 336)

Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman Jr., F.R., Muzny, D.

, Bouck, J., Gibbs, R.A. and Margolin, J.F.

Pediatric Leukemia CDNA Sequencing Project

Unpublished (2000)

Contact: Dr. Judith F. Margolin

Texas Children's Cancer Center and Human Genome Sequencing Center

at Baylor College of Medicine

1102 Bates, MC3-3320 Houston, TX 77030, USA

Tel: 832-824-4536

Fax: 832-825-4038

Email: clones@xccc.org

Citation: Carninci, P. and Hayashizaki, Y. High efficiency

full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Seq primer: M13 primer.

Location/Qualifiers

1. .336

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TCBAP5756"

/clone\_lib="Pediatric pre-B cell acute lymphoblastic

leukemia Baylor-HGSC project-TCBA"

/sex="male"

/tissue\_type="leukopheresis"

/cell\_type="pre-B cell"

/dev\_stage="pediatric 2 years"

/lab\_host="DH108"

/note="Vector: lambda PSB; Site\_1: BamHI; Site\_2: EcoRI;

First strand cDNA was primed with an anchored

XhoI-Oligo(dT) primer [5'GGAGACTCGAGCGCGCCGAGAGAG(T)VN

3'; V-A/C/G; N-A/C/G/T] and then dg tailed. Second strand

[5'AGAGCTCGATCCGCGCGCCGCAATATATAT(C) 3'].

Double-stranded cDNA was then digested with BamHI and XhoI

and directionally cloned into the BamHI and SalI sites of

lambda PSB vector. Library went through one round of

normalization. Library was constructed by Wei Yu at RIKEN

of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,

Itch M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,  
Schneider C, Hayashizaki Y. High efficiency selection of  
full-length cDNA by improved biotinylated cap trapper.,  
DNA Res 4: 1, 61-6, Feb 28, 1997)

BASE COUNT

99 a 43 c 56 g 138 t

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 336;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 ACAGTGATTAAATTAA 101

Db 31 ACAGTGATTAAATTAA 13

Search completed: October 27, 2002, 22:36:47  
Job time: 1544.65 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 20:19:49 : Search time 57.0363 Seconds  
(without alignments)  
3307.484 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768

Sequence: 1 atgtcagttgaattctacaa.....aggtgattagcaccgaaaaa 768

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Issued Patents\_NA:\*
- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
  - 3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
  - 4: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
  - 5: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
  - 6: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	255	33.2	486	1 US-08-048-164A-1	Sequence 1, Appl 1
2	255	33.2	486	1 US-08-048-164A-3	Sequence 3, Appl 1
3	255	33.2	486	1 US-08-460-462-1	Sequence 1, Appl 1
4	255	33.2	486	1 US-08-460-462-3	Sequence 3, Appl 1
5	255	33.2	486	1 US-08-460-457-1	Sequence 1, Appl 1
6	255	33.2	486	1 US-08-460-457-3	Sequence 3, Appl 1
7	255	33.2	486	1 US-08-460-458-1	Sequence 1, Appl 1
8	255	33.2	486	1 US-08-460-458-3	Sequence 3, Appl 1
9	255	33.2	486	2 US-08-460-455-1	Sequence 1, Appl 1
10	255	33.2	486	2 US-08-460-455-3	Sequence 3, Appl 1
11	255	33.2	486	2 US-08-330-394A-1	Sequence 1, Appl 1
12	255	33.2	486	2 US-08-330-394A-3	Sequence 3, Appl 1
13	255	33.2	486	2 US-09-006-636-5	Sequence 5, Appl 1
14	255	33.2	486	2 US-09-006-636-5	Sequence 5, Appl 1
15	255	33.2	486	2 US-09-277-716-21	Sequence 21, Appl 1
16	255	33.2	486	2 US-08-330-394A-23	Sequence 23, Appl 1
17	255	33.2	486	2 US-09-198-956-9	Sequence 9, Appl 1
18	255	33.2	486	2 US-09-198-955A-11	Sequence 11, Appl 1
19	255	33.2	486	2 US-08-916-917-3	Sequence 3, Appl 1
20	255	33.2	486	2 US-08-916-917-3	Sequence 3, Appl 1
21	255	33.2	486	2 US-08-972-631-3	Sequence 3, Appl 1
22	255	33.2	486	2 US-08-972-629-3	Sequence 3, Appl 1
23	255	33.2	486	2 US-08-972-630-3	Sequence 3, Appl 1
24	255	33.2	486	2 US-08-672-211-3	Sequence 3, Appl 1
25	255	33.2	486	2 US-09-225-170-3	Sequence 3, Appl 1
26	255	33.2	486	2 US-08-074-879-9	Sequence 9, Appl 1
27	255	33.2	486	2 US-08-468-057A-9	Sequence 9, Appl 1

c 28	17	2.2	754	4 US-09-541-941B-26	Sequence 26, Appl 1
c 29	17	2.2	1384	1 US-08-553-888A-1	Sequence 1, Appl 1
c 30	17	2.2	1466	1 US-08-553-888A-2	Sequence 2, Appl 1
c 31	17	2.2	4982	3 US-08-599-103B-1	Sequence 1, Appl 1
c 32	17	2.2	4982	4 US-09-229-059-1	Sequence 4, Appl 1
c 33	16	2.1	88	4 US-09-042-071-49	Sequence 49, Appl 1
c 34	16	2.1	707	2 US-08-465-380-37	Sequence 37, Appl 1
c 35	16	2.1	707	2 US-08-465-397-37	Sequence 37, Appl 1
c 36	16	2.1	707	2 US-08-466-399-37	Sequence 37, Appl 1
c 37	16	2.1	707	2 US-08-461-965-37	Sequence 37, Appl 1
c 38	16	2.1	707	2 US-08-634-641-37	Sequence 37, Appl 1
c 39	16	2.1	707	3 US-09-249-471-37	Sequence 37, Appl 1
c 40	16	2.1	707	3 US-09-249-472-37	Sequence 37, Appl 1
c 41	16	2.1	707	3 US-09-249-451-37	Sequence 37, Appl 1
c 42	16	2.1	707	3 US-08-809-455-37	Sequence 37, Appl 1
c 43	16	2.1	707	3 US-09-249-461-37	Sequence 37, Appl 1
c 44	16	2.1	707	3 US-09-249-448-37	Sequence 37, Appl 1
c 45	16	2.1	787	4 US-08-858-207A-210	Sequence 210, Appl 1

ALIGNMENTS

RESULT 1  
US-08-048-164A-1  
Sequence 1, Application US/08048164A  
Patent No. 5496934  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shlegel, Itai  
APPLICANT: Goldstein, Marc A.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/048,164A  
FILING DATE: 14-Apr-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REFERENCE/DOCKET NUMBER: 18,872  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-048-164A-1  
Query Match 33.2%; Score 255; DB 1; Length 486;  
Best Local Similarity 100.0%; Pred. No. 5,5e-124;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	16	TACAACCTCTAAACAATCAGCAACAACAACTCAATTCCACCATAATCAAAATTTACTAAC	75
Db	31	TACAACCTCTAAACAATCAGCAACAACAACTCAATTCCACCATAATCAAAATTTACTAAC	90
Oy	76	ACATCTGCAGTGAATTAATTTAAATGACGTAAACGTATGATATTTATTCACAGAATGAT	135
Db	91	ACATCTGCAGTGAATTTAAATTTAAATGACGTAAACGTATGATATTTATTCACAGAATGAT	150
Oy	136	GGTACACAGGACAAACCTTTCTGCTGTGACCATCTGGTGCATTATTTAGGAATAGCTAT	195
Db	151	GGTACACAGGACAAACCTTTCTGCTGTGACCATCTGGTGCATTATTTAGGAATAGCTAT	210
Oy	196	GTTGATTAACACTAGCAAAAGTCACAGCAAACTTCTTTAAGAAACAGCAAGCCCAACATCA	255
Db	211	GTTGATTAACACTAGCAAAAGTCACAGCAAACTTCTTTAAGAAACAGCAAGCCCAACATCA	270
Oy	256	ACCTATGATACATAT 270	
Db	271	ACCTATGATACATAT 285	

```

1      RESULT 2
2      US-08-048-164A-3/C
3      Sequence 3, Application US/08048164A
4      Patent No. 5496934
5      GENERAL INFORMATION:
6      APPLICANT: Shoseyov, Oded
7      APPLICANT: Shpiegl, Itai
8      APPLICANT: Goldstein, Marc A.
9      APPLICANT: DOL, Roy H.
10     TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
11     NUMBER OF SEQUENCES: 21
12     CORRESPONDENCE ADDRESS:
13     ADDRESSEE: PENNIE & EDMONDS
14     STREET: 1155 Avenue of the Americas
15     CITY: New York
16     STATE: New York
17     COUNTRY: U.S.A.
18     ZIP: 10036
19     COMPUTER READABLE FORM:
20     MEDIUM TYPE: Floppy disk
21     COMPUTER: IBM PC compatible
22     OPERATING SYSTEM: PC-DOS/MS-DOS
23     SOFTWARE: Patentln Release #1.0, Version #1.25
24     CURRENT APPLICATION DATA:
25     APPLICATION NUMBER: US/08/048,164A
26     FILING DATE: 14-APR-1993
27     CLASSIFICATION: 435
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Mistock, S. Leslie
30     REGISTRATION NUMBER: 18,872
31     REFERENCE/DOCKET NUMBER: 7809-003
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (212) 790-9090
34     TELEFAX: (212) 869-8864/9741
35     TELEX: 66141 PENNIE
36     INFORMATION FOR SEQ ID NO: 3:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 486 base pairs
39     TYPE: nucleic acid
40     STRANDEDNESS: double
41     TOPOLOGY: unknown
42     MOLECULE TYPE: DNA
43     US-08-048-164A-3

```

Query Match	33.28;	Score 255;	DB 1;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 5.3e-124;		
Matches 255; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 16 TACAACCTTACAACATGACACAAACAACTCAAATTACACCAATTAATCAAAATTACTTAAC 75  
|||||  
D6 456 TACAACCTTACAACATGACACAAACAACTCAAATTACACCAATTAATCAAAATTACTTAAC 397TAC

QY	76	ACATCTGCAGTGGATTTAAATTTAAATGACGTAAAGTTAGATTTTTCACACAAGTAT	135
Db	396	ACATCTGCAGTGGATTTAAATTTAAATGACGTAAAGTTAGATTTTTCACACAAGTAT	357
QY	136	GSTACACAAAGACAAACCTTTTGGTGACCATGCTGTGCATATATAGAGAAATAGCTAT	195
Db	336	GSTACACAAAGACAAACCTTTTGGTGACCATGCTGTGCATATATAGAGAAATAGCTAT	277
QY	196	GTTGATTAACACTAGCAAAAGTGACACGAAACCTTCGTTAAAGAACAGCAAGCCCAACATCA	255
Db	276	GTTGATTAACACTAGCAAAAGTGACACGAAACCTTCGTTAAAGAACAGCAAGCCCAACATCA	217
QY	256	ACCTATGATACATAT 270	
Db	216	ACCTATGATACATAT 202	

```

1      RESULT 3
2      US-08-460-462-1
3      / Sequence 1, Application US/08460462
4      / Patent No. 5670623
5      / GENERAL INFORMATION:
6      / APPLICANT: Shoseyov, Oded
7      / APPLICANT: Shpiegl, Itai
8      / APPLICANT: Goldstein, Marc A.
9      / APPLICANT: Doi, Roy H.
10     / TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
11     / NUMBER OF SEQUENCES: 21
12     / CORRESPONDENCE ADDRESS:
13     / ADDRESSEE: PENNIE & EDMONDS
14     / STREET: 1155 Avenue of the Americas
15     / CITY: New York
16     / STATE: New York
17     / COUNTRY: U.S.A.
18     / ZIP: 10036
19     / COMPUTER READABLE FORM:
20     / MEDIUM TYPE: Floppy disk
21     / COMPUTER: IBM PC compatible
22     / OPERATING SYSTEM: PC-DOS/MS-DOS
23     / SOFTWARE: PatentIn Release #1.0, Version #1.25
24     / CURRENT APPLICATION DATA:
25     / APPLICATION NUMBER: US/08/460,462
26     / FILING DATE: concurrently herewith
27     / CLASSIFICATION: 435
28     / PRIOR APPLICATION DATA:
29     / APPLICATION NUMBER: US 08/048,164
30     / FILING DATE: 14-APR-1993
31     / ATTORNEY/AGENT INFORMATION:
32     / NAME: Mistrock, S. Leslie
33     / REGISTRATION NUMBER: 16,872
34     / REFERENCE/DOCKET NUMBER: 7809-006
35     / TELECOMMUNICATION INFORMATION:
36     / TELEPHONE: (212) 790-9090
37     / TELEFAX: (212) 869-8864/9741
38     / TELEX: 66141 PENNIE
39     / INFORMATION FOR SEQ ID NO: 1:
40     / SEQUENCE CHARACTERISTICS:
41     / LENGTH: 486 base pairs
42     / TYPE: nucleic acid
43     / STRANDEDNESS: double
44     / TOPOLOGY: unknown
45     / MOLECULE TYPE: DNA
46     / FEATURE:
47     / NAME/KEY: CDS
48     / LOCATION: 1..486
49     /
50     / US-08-460-462-1

```

Query Match	33.28;	Score 255;	DB 1;	Length 486;
Best Local Similarity	100.0%;	Pred. No. 5.5e-124;		
Matches 255; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

16 TACAACCTTAACAATCAGCACAACAACCTCAATTACACCAATATCAAAATTACTAAC 75

|||||  
Db 31 TACAACCTTAACAANTGACGACAAACAACTCAATTACACCAATAATCAAAATTACTAAC 90  
Qy 76 ACATCTGACAGTATTAATTAATTAATGAGTAAAGTTGATATTTTACACAGGAT 135  
Db 91 ACATCTGACAGTATTAATTAATTAATGAGTAAAGTTGATATTTTACACAGGAT 150  
Qy 136 GTTACACAAAGGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 195  
Db 151 GTTACACAAAGGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 210  
Qy 196 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAAACAGCAAGCCCAATCA 255  
Db 211 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAAACAGCAAGCCCAATCA 270  
Qy 256 ACCTATGATACATAT 270  
Db 271 ACCTATGATACATAT 285

## RESULT 4

US-08-460-462-3/C  
Sequence 3, Application US/08460462  
Patent No. 5670623  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shlegel, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,462  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-460-462-3

Query Match 33.2%; Score 255; DB 1; Length 486;  
Best Local Similarity 100.0%; Pred. No. 5,5e-124;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 16 TACAACCTTAACAANTGACGACAAACAACTCAATTACACCAATAATCAAAATTACTAAC 75  
|||||

Db 456 TACAACCTTAACAANTGACGACAAACAACTCAATTACACCAATAATCAAAATTACTAAC 397  
Qy 76 ACATCTGACAGTATTAATTAATTAATGAGTAAAGTTGATATTTTACACAGGAT 135  
Db 396 ACATCTGACAGTATTAATTAATTAATGAGTAAAGTTGATATTTTACACAGGAT 337  
Qy 136 GTTACACAAAGGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 195  
Db 336 GTTACACAAAGGACAACTTCTGCTGACCAATGCTGTCATTTATAGAAATAGCTAT 277  
Qy 196 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAAACAGCAAGCCCAATCA 255  
Db 276 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAAACAGCAAGCCCAATCA 217  
Qy 256 ACCTATGATACATAT 270  
Db 216 ACCTATGATACATAT 202

## RESULT 5

US-08-460-457-1  
Sequence 1, Application US/08460457  
Patent No. 5719044  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shlegel, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,457  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-457-1

Query Match 33.2%; Score 255; DB 1; Length 486;  
Best Local Similarity 100.0%; Pred. No. 5,5e-124;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	16	TACAACTTACCAATGAGCAACAACGATTTACCACAATTAATCAAAATTACTAAC	75
Db	31	TACAACTTAACAAATGAGCAACAACAACGATTTACCACAATTAATCAAAATTACTAAC	90
OY	76	ACATCTGCAGTAGTTTAAATTTAAATGAGCTAAAGTTAGATTTATTCACAGTGTAT	135
Db	91	ACATCTGCAGTAGTTTAAATTTAAATGAGCTAAAGTTAGATTTATTCACAGTGTAT	150
OY	136	GSTACACAGAGCAACAACCTTTGGTGTGACACATCTGTCATATTATGAGAAATAGTAT	195
Db	151	GSTACACAGAGCAACAACCTTTGGTGTGACATCTGTCATATTATGAGAAATAGTAT	210
OY	196	GTTGATAACACTAGCAAGTAGACAGCAAACTTGTTAAAGAAACGACAAAGCCCAACATCA	255
Db	211	GTTGATAACACTAGCAAGTAGACAGCAAACTTGTTAAAGAAACGACAAAGCCCAACATCA	270
OY	256	ACCTATGATACATAT 270	?
Db	271	ACCTATGATACATAT 285	

```

RESULT 6
US-08-460-457-3/C
Sequence 3, Application US/08460457
Patent No. 5719044
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Dol, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,457
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-460-457-3

Query Match 33.2%; Score 255; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 5.5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
16 TACAACTCTAACCAATTCAGCACAACAACCAATTCATTCACCAATATATCAAAATTACTAC 75

```

Db	456	TACAACCTCTAGCAACATCGACCAACAACGTCATTTACACCATTAATCAAAATTACTAAC	397
Oy	76	ACATCTGCACGTGATTTTAAATTTTAAAGAGCTAAACGTTAGATTTTACACAAGGAT	135
Db	336	ACATCTGCACGTGATTTTAAATTTTAAAGAGCTAAAGTTGATATTTATACACAAGTAT	357
Oy	136	GGTACACAGAGACAACACTTTCTGCTGTGACCACTGCTGTCATTTATTAGGAATAGCTAT	195
Db	336	GGTACACAGAGACAACACTTTCTGCTGTGACCACTGCTGTCATTTATTAGGAATAGCTAT	277
Oy	196	GTTGATTAACACTTACCAAAAGTACACGCAAACTTCGTTAAAGAAACAGCAACCCCAACATCA	255
Db	276	GTTGATTAACACTTACCAAAAGTACACGCAAACTTCGTTAAAGAAACAGCAACCCCAACATCA	217
Oy	256	ACCTATGATACATAT	270
Db	216	ACCTATGATACATAT	202

```

RESULT 7
US-08-460-458-1
: Sequence 1, Application US/08460458
: Patent No. 5738984
: GENERAL INFORMATION:
: APPLICANT: Shoseyov, Oded
: TITLE OF INVENTION: 'KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAINS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,458
: FILING DATE: concurrently herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,164
: FILING DATE: 14-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7809-007
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 486 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..486
:
: US-08-460-458-1
:
: Query Match 33.2%, Score 255; DB 1; Length 486;
: Best Local Similarity 100.0%; Pred. No. 5.5e-124;
: Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: 16 TACAACCTTACAATTCAGCACAAACAATCAATTAACCAATTAATCAAAATTACTAAC 75
: |||||||

```



Db 31 TACAACTCTAACAAATCAGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 90  
QY 76 ACATCTGACAGTATTAATTAATGAGTAAAGTTAGATATTTTACACAAGTAT 135  
Db 91 ACATCTGACAGTATTAATTAATGAGTAAAGTTAGATATTTTACACAAGTAT 150  
QY 136 GGTATACAAAGACAACATTTCTGTGTGACCATCTGTCATTTATAGGAAATAGCTAT 195  
Db 151 GGTACACAAAGACAACATTTCTGTGTGACCATCTGTCATTTATAGGAAATAGCTAT 210  
QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGTTTAAAGAACAGCAAGCCCAACATCA 255  
Db 211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGTTTAAAGAACAGCAAGCCCAACATCA 270  
QY 256 ACCTATGATACATAT 270  
Db 271 ACCTATGATACATAT 285

## RESULT 8

US-08-460-458-3/C  
Sequence 3, Application US/08460458  
Patent No. 5738984  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,458  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-460-458-3

Query Match 33.2%; Score 255; DB 1; Length 486;  
Best Local Similarity 100.0%; Pred. No. 5,5e-124;

Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 16 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATATCAAAATTAAGTAC 75  
Db 456 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATATCAAAATTAAGTAC 397  
QY 76 ACATCTGACAGTATTAATTAATGAGTAAAGTTAGATATTTTACACAAGTAT 135

Db 396 ACATCTGACAGTATTAATTAATGAGTAAAGTTAGATATTTTACACAAGTAT 337  
QY 136 GGTACACAAAGACAACATTTCTGTGTGACCATCTGTCATTTATAGGAAATAGCTAT 195  
Db 336 GGTACACAAAGACAACATTTCTGTGTGACCATCTGTCATTTATAGGAAATAGCTAT 277  
QY 196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGTTTAAAGAACAGCAAGCCCAACATCA 255  
Db 276 GTTGATTAACACTAGCAAAAGTGACAGCAAACTGTTTAAAGAACAGCAAGCCCAACATCA 217  
QY 256 ACCTATGATACATAT 270  
Db 216 ACCTATGATACATAT 202

## RESULT 9

US-08-460-455-1  
Sequence 1, Application US/08460455  
Patent No. 5837814  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Doi, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,455  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7809-009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..486  
US-08-460-455-1

Query Match 33.2%; Score 255; DB 2; Length 486;  
Best Local Similarity 100.0%; Pred. No. 5,5e-124;

Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 16 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATATCAAAATTAAGTAC 75  
Db 31 TACAACCTTAACAATGACGACAAACAACCTCAATTACACCAATATCAAAATTAAGTAC 90



Db 96 ACATCTGCACGTGATTTAAATTTAAATGACGTAAAGTTGATTTTATTCACACAGTGAT 155

```

Db 156 GGTACACAGAGCAAACTTCTGCTGACCATCTGTCATATTAGGAATAGCTAT 215
|
QY 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 255
|
Db 216 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 275
|
QY 256 ACCTATGATACATAT 270
|
Db 276 ACCTATGATACATAT 290
|
RESULT 14
US-09-006-632-5
; Sequence 5, Application US/09006632
; Patent No. 6184440
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shani, Ziv
; APPLICANT: Shpigel, Etai
; TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED
; MORPHOLOGY
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,632
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7809-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 499 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-006-632-5
Query Match 33.2%; Score 255; DB 4; Length 499;
Best Local Similarity 100.0%; Pred. No. 5.5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 TACAACCTCTAACAATGAGCAACAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
|
Db 36 TACAACCTCTAACAATGAGCAACAACAACTCAATTACACCAATTAATCAAAATTACTAAC 95
|
QY 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135
|
Db 96 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 155
|
QY 136 GGTACACAGAGCAAACTTCTGCTGACCATCTGTCATATTAGGAATAGCTAT 195
|
Db 156 GGTACACAGAGCAAACTTCTGCTGACCATCTGTCATATTAGGAATAGCTAT 215
|
QY 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 255
|
Db 216 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 275
|
QY 256 ACCTATGATACATAT 270
|
Db 276 ACCTATGATACATAT 290
|

```

```

Db 216 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 275
|
QY 256 ACCTATGATACATAT 270
|
Db 276 ACCTATGATACATAT 290
|
RESULT 15
US-09-277-716-21
; Sequence 21, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; FEATURE:
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gausia lucifera
; US-09-277-716-21
Query Match 33.2%; Score 255; DB 4; Length 1146;
Best Local Similarity 100.0%; Pred. No. 5.5e-124;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 TACAACCTCTAACAATGAGCAACAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
|
Db 16 TACAACCTCTAACAATGAGCAACAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
|
QY 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135
|
Db 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135
|
QY 136 GGTACACAGAGCAAACTTCTGCTGACCATCTGTCATATTAGGAATAGCTAT 195
|
Db 136 GGTACACAGAGCAAACTTCTGCTGACCATCTGTCATATTAGGAATAGCTAT 195
|
QY 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 255
|
Db 196 GTTGATTAACACTAGACAGAGCAAACTGCTTAAGAAAGAGCCCAACTCA 255
|
QY 256 ACCTATGATACATAT 270
|
Db 256 ACCTATGATACATAT 270
|
Search completed: October 27, 2002, 23:22:28
Job time : 63.0363 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic acid - nucleic acid search, using sw model

Run on: October 27, 2002, 17:46:54 : Search time 194.542 Seconds  
(without alignments)  
6777.910 Million cell updates/sec

Title: US-09-677-374-5

Sequence: 1 atgtcagttgaattctacaa.....agggtgattagcaccggaanaa 768

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: N.Geneseq-032802.\*

```
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	100.0	768	22	AAF86248
2	483	62.9	483	22	AAF86247
3	270	35.2	573	22	AD11043
4	270	35.2	1030	22	AD11044
5	255	33.2	486	15	AA072917
6	255	33.2	486	20	AA724072
7	255	33.2	489	20	AA244930
8	255	33.2	507	22	AD11042
9	255	33.2	984	22	AD11046

10	255	33.2	1146	20	AA227550	Gaussia luciferase
11	255	33.2	1146	24	AA222301	Gaussia species CB
12	255	33.2	1288	22	AD11045	Clostridium cellu
13	118	15.4	118	22	AAF86254	PCR primer #4 used
14	104	13.5	110	22	AAF86252	PCR primer #2 used
15	102	13.3	102	22	AAF86255	PCR primer #5 used
16	87	11.3	111	22	AAF86251	PCR primer #1 used
17	75	9.8	94	22	AAF86253	PCR primer #3 used
18	40	5.2	110	22	AAF86256	PCR primer #6 used
19	34	4.4	38	21	AA427285	PCR primer for fru
20	20	2.6	26	20	AAV74089	Plasmid pET-R1-CBD
21	20	2.6	3833	23	ABLO2577	Drosophila melanog
22	19	2.5	808	19	AA286688	Ripening banana pu
23	19	2.5	892	22	AA407401	Human cDNA clone (
24	19	2.5	1438	21	AA45336	DNA encoding a man
25	19	2.5	1482	20	AA90978	DNA encoding pecta
26	19	2.5	1482	20	AA431562	pectate lyase CBD
27	19	2.5	2042	22	AA116496	Human cDNA sequen
28	19	2.5	2562	22	AA159154	Human polynucleoti
29	19	2.5	2735	22	AA327244	Human secreted pro
30	19	2.5	2878	22	AA502416	Human secreted pro
31	19	2.5	4283	22	AA525994	Human cDNA encodin
32	19	2.5	5562	19	AA786625	C. thermocellum C1
33	19	2.5	6223	22	AA158572	Human polynucleoti
34	19	2.5	6304	22	AA158574	Human polynucleoti
35	19	2.5	6382	22	AA158573	Human polynucleoti
36	18	2.3	1029	22	AA452017	Mycobacterium tube
37	18	2.3	1237	23	AA594517	DNA encoding novel
38	18	2.3	1300	18	AA766951	Asialloglycoprotein
39	18	2.3	1353	21	AA39312	Arabidopsis thalia
40	18	2.3	1356	23	AA559780	Propionibacterium
41	18	2.3	1445	23	AA577372	DNA encoding novel
42	18	2.3	1623	23	AA582377	DNA encoding novel
43	18	2.3	1903	22	AD11584	Human novel protei
44	18	2.3	3684	23	AA584339	DNA encoding novel
45	18	2.3	3741	22	AA81357	Escherichia coli p

#### ALIGNMENTS

RESULT 1	AAF86248	standard; DNA; 768 BP.
ID	AAF86248:	
XX	11-JUL-2001	(first entry)
XX	DNA sequence of c17e2 ospa construct with N-terminal fusion partner.	
DE	Poikilothermic fish; piscirickettsia salmonis; rickettsial pathogen;	
KW	vaccine; OSPA; salmonid rickettsial septicaemia; rickettsial disease;	
KW	SRS; 17E2; fusion construct; ds.	
XX	piscirickettsia salmonis.	
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..768
FT		/*tag- a
FT		/partial
FT		/product= "c17e2 Ospa with N-terminal fusion partner"
FT		/note= "No stop codon is given"
FT	misc_feature	1..285
FT		/*tag- b
FT		/note= "DNA encoding undefined N-terminal fusion partner"
FT	misc_feature	286..768
FT		/*tag- c
FT		/note= "Optimised Ospa construct c17e2"
XX	CA2281913-A1.	

```

PD 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZU/) KUZUK M A.
XX
XX Key WW, Burian J, Kuzuk MA.
XX
XX WPI: 2001-316844/34.
XX P-PSDB; AAB81128.
XX
XX Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the OspA protein of Piscirickettsia salmonis
XX
XX Example 4; Fig 5; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
XX of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
XX salmonis. The method comprises administering an immunogenic amount of a
XX P. salmonis specific antigen termed OspA, or an immunogenic fragment of
XX OspA in the form of a vaccine. The method is used for protecting animals,
XX particularly poikilothermic fish, against the bacterial pathogen
XX P. salmonis. The method is also useful for protecting against salmonid
XX rickettsial septicemia (SRS) and other rickettsial diseases. The present
XX sequence represents P. salmonis OspA DNA termed C17b2 optimised for
XX expression in Escherichia coli fused to DNA encoding an undefined
XX C-terminal fusion partner. The protein encoded by this fusion construct
XX is used in a vaccine to create an anti-OspA antibody response.
XX
XX Sequence 768 BP; 220 A; 183 C; 190 G; 175 T; 0 other:
XX
XX Query Match 100.0%; Score 768; DB 22; Length 768;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 768; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGCACTTGAATTTCTCAACTCTAACAAATCAGACAAACAACTCAATTACCAATA 60
XX |
XX 1 ATGCACTTGAATTTCTCAACTCTAACAAATCAGACAAACAACTCAATTACCAATA 60
XX
XX 61 ATCAAAATTTCTAACACATCTGACAGTGAATTAATTAATGAGTAAAGTAGATAT 120
XX |
XX 61 ATCAAAATTTCTAACACATCTGACAGTGAATTAATTAATGAGTAAAGTAGATAT 120
XX
XX 121 TATTACACAAGTGTGTACACAAGACAACCTTCTGTGTGACATGCTGTGATTA 180
XX |
XX 121 TATTACACAAGTGTGTACACAAGACAACCTTCTGTGTGACATGCTGTGATTA 180
XX
XX 181 TTAGGAATAGCTATGTTGATACACTAGCAAGTGCAGCAACTTCGTTAAAGAAACA 240
XX |
XX 181 TTAGGAATAGCTATGTTGATACACTAGCAAGTGCAGCAACTTCGTTAAAGAAACA 240
XX
XX 241 GCAAGCCCAACATCAACCTTATGATATCTGATCCGTCATGATGCGTGTGCTG 300
XX |
XX 241 GCAAGCCCAACATCAACCTTATGATATCTGATCCGTCATGATGCGTGTGCTG 300
XX
XX 241 GCAAGCCCAACATCAACCTTATGATATCTGATCCGTCATGATGCGTGTGCTG 300
XX
XX 301 CAGGCGAGCTCTGTGATCATATCTCTGTTCTGTGTGCGGCGCCAGAACTTACG 360
XX |
XX 301 CAGGCGAGCTCTGTGATCATATCTCTGTTCTGTGTGCGGCGCCAGAACTTACG 360
XX
XX 301 CAGGCGAGCTCTGTGATCATATCTCTGTTCTGTGTGCGGCGCCAGAACTTACG 360
XX
XX 361 CGCCAGAGAGTTGGCGCGGCGCACCGGTGCGTGTGCGCGCTGTTCCGCGCACCTGTC 420
XX |
XX 361 CGCCAGAGAGTTGGCGCGGCGCACCGGTGCGTGTGCGCGCTGTTCCGCGCACCTGTC 420
XX
XX 421 GGTAAGAGCTCTGTGCTGTGATGCGCATGCGGCGGTGCGGCTTGGGGGCGCTGATT 480
XX |
XX 421 GGTAAGAGCTCTGTGCTGTGATGCGCATGCGGCGGTGCGGCTTGGGGGCGCTGATT 480
XX
XX 481 GGCTCTAAATCGGTGACAGCATGAGCAGCAGAGATAAATCAAACTGAACCAAGTCTCTG 540

```

```

DB 481 GGCTCTAAATCGGTGACAGCATGAGCAGCAGAGATAAATCAAACTGAACCAAGTCTCTG 540
DB 541 GAAAGAGTGAAGCGCGCAGGTTACTGTTGGCGTATATCCGACACCGGTTAACAGCTTAC 600
DB 541 GAAAGAGTGAAGCGCGCAGGTTACTGTTGGCGTATATCCGACACCGGTTAACAGCTTAC 600
DB 601 TCTGTGGAACCGGTTGCGACCTTACACCGGTTACACAAACAGGACCGCTGACAGCTTAC 660
DB 601 TCTGTGGAACCGGTTGCGACCTTACACCGGTTACACAAACAGGACCGCTGACAGCTTAC 660
DB 661 TGCCGCAATTTTCAGCAGAAAGCCATGATGCAAGTTCAGAAACAGGAAATCTACGCGACC 720
DB 661 TGCCGCAATTTTCAGCAGAAAGCCATGATGCAAGTTCAGAAACAGGAAATCTACGCGACC 720
DB 721 GCGTGCCTTCAGCCGAGTGGCGCGTGCAGAGTATTTAGCACCCGAAAAA 768
DB 721 GCGTGCCTTCAGCCGAGTGGCGCGTGCAGAGTATTTAGCACCCGAAAAA 768

RESULT 2
AAF86247
ID AAF86247 standard; DNA; 483 BP.
XX
XX AAF86247;
AC
AC 11-JUL-2001 (first entry)
XX
XX DNA sequence of E. coli optimised ospa gene 17E2.
XX
XX Poikilothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
XX vaccine; OspA; salmonid rickettsial septicemia; rickettsial disease;
XX SRS; 17E2; ds.
XX
XX Piscirickettsia salmonis.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX 1..483
XX CDS
XX
XX CA2281913-A1.
XX
XX 17-MAR-2001.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX 17-SEP-1999; 99CA-2281913.
XX
XX (KAYW/) KAY W W.
XX (BURI/) BURIAN J.
XX (KUZU/) KUZUK M A.
XX
XX Key WW, Burian J, Kuzuk MA.
XX
XX WPI: 2001-316844/34.
XX P-PSDB; AAB81128.
XX
XX Method for protecting poikilothermic fish against salmonid rickettsial
XX septicemia and other rickettsial diseases comprises administering a
XX vaccine containing the OspA protein of Piscirickettsia salmonis
XX
XX Example 3; Fig 4C; 35pp; English.
XX
XX This invention relates to a method for the protection against infection
XX of a poikilothermic fish by the bacterial pathogen, Piscirickettsia
XX salmonis. The method comprises administering an immunogenic amount of a
XX P. salmonis specific antigen termed OspA, or an immunogenic fragment of
XX OspA in the form of a vaccine. The method is used for protecting animals,

```

CC particularly polikilothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents DNA which has been optimised for expression in  
 CC Escherichia coli to encode the P. salmonis OspA protein. An OspA protein  
 CC with an N-terminal fusion partner is used in a vaccine to create an  
 CC anti-OspA antibody response.

XX  
 SO Sequence 483 BP; 108 A; 127 C; 148 G; 100 T; 0 other;

Query Match 62.9%; Score 483; DB 22; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 6, 1e-242;

Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 ATGCGTGCTGCTGCGAGGAGCTCTGTATCATTTATCTGTGTTTCTGCGGTGC 345  
 DB 1 ATGCGGTGCTGCGAGGAGCTCTGTATCATTTATCTGTGTTTCTGCGGTGC 60  
 OY 346 GCCAGAACTTCAAGCCAGGAAGTTGGCGGCCACCGGTGGTGTGGCGGTGTT 405  
 DB 61 GCCAGAACTTCAAGCCAGGAAGTTGGCGGCCACCGGTGGTGTGGCGGTGTT 120  
 OY 406 GCCGCGACCTGTGCTGTAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 465  
 DB 121 GCCGCGACCTGTGCTGTAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180  
 OY 466 CTGGGCGGTGTGATGCTTAAATCGGTGACAGCATGACAGCAGATTAATATAA 525  
 DB 181 CTGGGCGGTGTGATGCTTAAATCGGTGACAGCATGACAGCAGATTAATATAA 240  
 OY 526 CTGACACAGTCTGTGAAAAAGTGAAGCGGCGAGTTCTGTTGGGTAAATCCGAC 585  
 DB 241 CTGACACAGTCTGTGAAAAAGTGAAGCGGCGAGTTCTGTTGGGTAAATCCGAC 300  
 OY 586 ACCGTAACAGTACTGTGTGAACCGGTTCGACACTACAGCGTTTCAACAAACAGAA 645  
 DB 301 ACCGTAACAGTACTGTGTGAACCGGTTCGACACTACAGCGTTTCAACAAACAGAA 360  
 OY 646 CGCGGTGACAGTACTGTGTGAACCGGTTCGACACTACAGCGTTTCAACAAACAGAA 705  
 DB 361 CGCGGTGACAGTACTGTGTGAACCGGTTCGACACTACAGCGTTTCAACAAACAGAA 420  
 OY 706 GAAATCTACGACCGCGGTGCTGACGCGATGCGGTGACGAGTATGACCGAA 765  
 DB 421 GAAATCTACGACCGCGGTGCTGACGCGATGCGGTGACGAGTATGACCGAA 480  
 OY 766 AAA 768  
 DB 481 AAA 483

RESULT 3

AAD11043 ID AAD11043 standard; DNA: 573 BP.

XX AAD11043;

XX 24-SEP-2001 (first entry)

DE Clostridium cellulovorans cellulose binding domain-180 (CBD-180) DNA.

XX polysaccharide modification; polysaccharide binding domain; PBD; paper;

KM yarn; fiber; textile; biological crosslinker; mechanical property;

KM Clostridium cellulovorans cellulose binding domain; wet strength;

KM durability; elasticity; CBDelos; cellulose binding protein A; CBP A;

XX CBD-180; ds.

OS Clostridium cellulovorans.

XX

PH Key

FT CDS

Location/Qualifiers

3..572

/tag- a

/product- "Cellulose binding domain-180 protein"

FT /note= "CDS does not include stop codon"

PN /partial

XX WO200134091-A2.

XX 17-MAY-2001.

XX 02-NOV-2000; 2000MO-1100708.

XX 08-NOV-1999; 99US-0164140.

XX 18-NOV-1999; 99US-0166389.

XX (CBD- ) CBD TECHNOLOGIES LTD.

XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

XX Levy I, Nussinovitch A, Shoseyov O;

XX MPI: 2001-457121/49.

XX P-PsDB; AAE05746.

XX Preparation of a polysaccharide containing material having at least one

XX desired structural, chemical, physical, electrical and/or mechanical

XX property

XX Example 1.2; Fig 1e-1g; 121pp; English.

XX The present invention relates to methods and compositions for cross-

XX linking and/or modifying the properties of polysaccharide materials.

XX The method involves treating the polysaccharide structure with a

XX polysaccharide binding domain (PBD) fusion protein. The method is

XX used to alter the structural, chemical, physical, electrical and

XX mechanical properties of polysaccharide materials such as paper,

XX yarns, fibers and textiles, using biological crosslinking agents.

XX The polysaccharide containing materials have improved mechanical

XX properties such as wet strengths, durability and elasticity. The PBD

XX reagent is applied in the forming stage in fusing paper manufacture

XX which eliminates the sizing step. The use of a biological crosslinker

XX improves the recyclability of paper products. The PBD reagent maintains

XX the fine fibers in a slurry therefore resulting in better recovery of

XX raw materials. The PBD molecules are eluted by strong alkaline conditions

XX which enhances the ability of the alkaline glue used in binding to

XX penetrate paper.

XX The present sequence is a DNA encoding Clostridium cellulovorans

XX cellulose binding domain (CBD)-180. CBD-180 contains 180 amino acids of

XX Clostridium cellulovorans CBD (CBDelos) of cellulose binding protein A

XX (CBP A).

SO Sequence 573 BP; 215 A; 112 C; 98 G; 148 T; 0 other;

Query Match 35.2%; Score 270; DB 22; Length 573;

Best Local Similarity 100.0%; Pred. No. 1e-130;

Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGTCAGTGAATCTCAACTCTAACAAATCAAGCAACAAACATCAATTAACACATA 60  
 DB 3 ATGTCAGTGAATCTCAACTCTAACAAATCAAGCAACAAACATCAATTAACACATA 62  
 OY 61 ATCAAAATTAACACATCTGACAGTGAATTAATTAATGACGTAAGTTAGTAT 120  
 DB 63 ATCAAAATTAACACATCTGACAGTGAATTAATTAATGACGTAAGTTAGTAT 122  
 OY 121 TATTACACAGTATGTTACACAAAGCAAACTTCTGGGTGACCACTGCTGTCATTA 180  
 DB 123 TATTACACAGTATGTTACACAAAGCAAACTTCTGGGTGACCACTGCTGTCATTA 182  
 OY 181 TTAGGAATAGTATGTTGTAACACTAGCAAGTGACAGCAAACTTCTGTTAAGAACA 240  
 DB 183 TTAGGAATAGTATGTTGTAACACTAGCAAGTGACAGCAAACTTCTGTTAAGAACA 242  
 OY 241 GCAAGCCCAACATCACTTATGATACATAT 270  
 DB 243 GCAAGCCCAACATCACTTATGATACATAT 272

```

RESULT 4
AADI1044
ID AADI1044 standard; DNA: 1030 BP.
XX
AC AADI1044;
XX
DT 24-SEP-2001 (first entry)
XX
DE Clostridium cellulovorans CBD cross linker protein (CCP) encoding DNA.
XX
KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW yarn; fiber; textile; biological crosslinker; mechanical property;
KW wet strength; durability; elasticity; cellulose binding domain; CBD;
KW CBD cross linker protein; CCP; ds.
XX
OS Clostridium cellulovorans.
XX
FH Key Location/Qualifiers
FT CDS 3..1028
FT /tag= a
FT /product= "CCP protein"
FT /transl_except= "(pos: 1020..1028, aa:Pro-Asp)"
FT /note= "CDS does not include stop codon"
FT /partial
XX
PN WO200134091-A2.
XX
PD 17-MAY-2001.
XX
PF 02-NOV-2000; 2000WO-IL00708.
XX
PR 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX
PA (CBD-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
PI Levy I, Nussinovitch A, Shoseyov O;
XX
DR MPI: 2001-457121/49.
DR P-PSDB; AAE05747.
XX
PT Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX
PS Example 1.2; Fig 2b-2e; 121pp; English.
XX
CC The present invention relates to methods and compositions for cross-
CC linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in futing, paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is a DNA encoding Clostridium cellulovorans
CC cellulose binding domain (CBD) cross linker protein (CCP).
XX
SO Sequence 1030 BP; 402 A; 195 C; 166 G; 267 T; 0 other;
Query Match 35.2%; Score 270; DB 22; Length 1030;
Best Local Similarity 100.0%; Pred. No. 1e-130;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ATGTCAGTTGAATCTACACCTTACAAATTCAGCACAACAACTCAATTACACCATATA
DB 3 ATGTCAGTTGAATCTACACCTTACAAATTCAGCACAACAACTCAATTACACCATATA
OY 61 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGAGTAAAGTTAGATAT
DB 63 ATCAAAATTTACTAACACATCTGACAGTGATTTAAATTTAAATGAGTAAAGTTAGATAT
OY 121 TATTACACAGTGTGATGACACAGACAACTTTGCTGCTGACATGTCATTA 180
DB 123 TATTACACAGTGTGATGACACAGACAACTTTGCTGCTGACATGTCATTA 182
OY 181 TTAGGAATTTAGCTATGTTGATTAACACTAGCAAAAGTGCACGAACTGCTTAAGAACA 240
DB 183 TTAGGAATTTAGCTATGTTGATTAACACTAGCAAAAGTGCACGAACTGCTTAAGAACA 242
OY 241 GCAAGCCCAACATCAACCTTATGATACATAT 270
DB 243 GCAAGCCCAACATCAACCTTATGATACATAT 272
RESULT 5
AAQ72917
ID AAQ72917 standard; DNA: 486 BP.
XX
AC AAQ72917;
XX
DT 05-JUN-1995 (first entry)
XX
DE Cellulose binding domain.
XX
KW Cellulose binding domain; CBD; ds.
XX
OS Clostridium cellulovorans.
XX
FH Key Location/Qualifiers
FT CDS 1..486
FT /tag= a
XX
PN WO9424158-A.
XX
PD 27-OCT-1994.
XX
PF 14-APR-1994; 94WO-US04132.
XX
PR 14-APR-1993; 93US-0048164.
XX
PA (REGC ) UNIV CALIFORNIA.
PA (YISS ) YISSUM RES & DEV CO.
PI Doi RH, Goldstein MA, Shoseyov O, Shpiegel I;
XX
DR MPI: 1994-341767/42.
DR P-PSDB; AAR62634.
XX
PT Isolated cellulose binding domain and fusion proteins - with
PT applications, including drug delivery, affinity separations, and
PT diagnostic techniques
XX
PS Claim 10; Fig 1; 125pp; English.
XX
CC AAQ72917/R63634 is a novel isolated cellulose binding domain. It pref.
CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,
CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and
CC the encoding DNA are claimed, as is a CBD fusion protein (FP)
CC comprising the CBD and a second protein. The second protein is pref.
CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive
CC HSP-related protein or an antigenic portion of this. The CBD and FP
CC may be used in drug delivery, affinity seps. and diagnostic
CC techniques. CBD nucleic acid may be obtd. from a variety of cell
CC sources that produce CBDs that bind with high affinity and in a

```



CC reversible manner or that produce CBD encoding mRNA. The preferred  
 CC source of CBD encoding nucleic acid is C. cellulovorans.  
 XX  
 SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 33.2%; Score 255; DB 15; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 7e-123;  
 Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACAAACAACCAATTCACACCAATATCAAAATTACTAAC 75  
 DB 31 TACAACCTTAACAATGACGACGACAAACAACCAATTCACACCAATATCAAAATTACTAAC 90  
 QY 76 ACATCTGACAGTGATTTAAATTTAATGAGCTAAAGTTAGATTATTTACACAGAGGAT 135  
 DB 91 ACATCTGACAGTGATTTAAATTTAATGAGCTAAAGTTAGATTATTTACACAGAGGAT 150  
 QY 136 GGTACACAGGACAACTTCTGTGTGACCATCTGTGCATTTATAGCAAAATAGCTAT 195  
 DB 151 GGTACACAGGACAACTTCTGTGTGACCATCTGTGCATTTATAGCAAAATAGCTAT 210  
 QY 196 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAACGCAAGCCCAACATCA 255  
 DB 211 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAACGCAAGCCCAACATCA 270  
 QY 256 ACCTATGATACATAT 270  
 DB 271 ACCTATGATACATAT 285

RESULT 6  
 AAV74072  
 ID AAV74072 standard; DNA: 486 BP.  
 XX  
 AC AAV74072;  
 XX  
 DT 09-MAR-1999 (first entry)  
 XX  
 DE C. cellulovorans cbpa CBD DNA.

XX  
 KW Cellulose binding domain; CBD; cbpa; fusion protein; diagnosis; HSP;  
 KW Immunoassay; heat-shock protein; cross reactive protein; detection;  
 KW antigenic fragment; antibody; insulin-dependent diabetes mellitus;  
 KW cellulose; chitin; cellulolytic; amorphogenic; ds.  
 XX  
 OS Clostridium cellulovorans.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1..486  
 FT /tag- a  
 FT /product- "cellulose binding domain"

XX  
 PN US5856201-A.  
 PD 05-JAN-1999.  
 XX  
 PE 27-OCT-1994; 94US-0330394.  
 XX  
 PR 27-OCT-1994; 94US-0330394.  
 PR 14-APR-1993; 93US-0048164.  
 PR 14-APR-1994; 94WO-US04132.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 PA (YISS ) YISSUM RES & DEV CO.  
 XX  
 PI Del RH, Goldstein MA, Shoseyov O, Shpiegl I;  
 DR P-PSDB; AAW90077.  
 XX  
 DR WPI: 1999-105130/09.  
 XX  
 PT Detection of a specific analyte by reaction with binding agent fused  
 PT to cellulose binding domain - and subsequent treatment with  
 PT cellulose and reaction of insoluble product with a label specific

PT for the analyte  
 XX  
 XX Example 7.2.1; Fig 1A-B; 63pp; English.

CC This sequence encodes a cellulose binding domain (CBD) derived from the  
 CC cbpa protein of Clostridium cellulovorans. The sequence is used the  
 CC construction of a fusion protein which can be used in diagnostic  
 CC immunoassays, e.g. to detect heat-shock proteins (HSP) and their  
 CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies  
 CC (which indicate insulin-dependent diabetes mellitus, or susceptibility  
 CC to it). The CBD, has very high affinity for cellulose (including  
 CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but  
 CC has almost no cellulolytic or amorphogenic activities. The CBD binds over  
 CC a wide pH range and is not released from cellulose by washing with water.  
 XX  
 SQ Sequence 486 BP; 194 A; 89 C; 79 G; 124 T; 0 other;

Query Match 33.2%; Score 255; DB 20; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 7e-123;  
 Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACGACGACAAACAACCAATTCACACCAATATCAAAATTACTAAC 75  
 DB 31 TACAACCTTAACAATGACGACGACGACAAACAACCAATTCACACCAATATCAAAATTACTAAC 90  
 QY 76 ACATCTGACAGTGATTTAAATTTAATGAGCTAAAGTTAGATTATTTACACAGAGGAT 135  
 DB 91 ACATCTGACAGTGATTTAAATTTAATGAGCTAAAGTTAGATTATTTACACAGAGGAT 150  
 QY 136 GGTACACAGGACAACTTCTGTGTGACCATCTGTGCATTTATAGCAAAATAGCTAT 195  
 DB 151 GGTACACAGGACAACTTCTGTGTGACCATCTGTGCATTTATAGCAAAATAGCTAT 210  
 QY 196 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAACGCAAGCCCAACATCA 255  
 DB 211 GTTGATTAACACTAGCAAAAGTGACGACAACTTCTGTTAAAGAACGCAAGCCCAACATCA 270  
 QY 256 ACCTATGATACATAT 270  
 DB 271 ACCTATGATACATAT 285

RESULT 7  
 AAX24930  
 ID AAX24930 standard; DNA: 499 BP.  
 XX  
 AC AAX24930;  
 XX  
 DT 21-JUN-1999 (first entry)  
 XX  
 DE Clostridium cellulovorans cbpa cellulose binding domain DNA.

XX  
 KW Cellulose binding domain; CBD; cbpa; endo-1,4-beta-glucanase; Cell;  
 KW Arabidopsis thaliana; transgenic plant; crop improvement;  
 KW morphology; cell wall; ds.  
 XX  
 OS Clostridium cellulovorans.

XX  
 PN WO9907830-A1.  
 PD 18-FEB-1999.  
 XX  
 PE 26-JUL-1998; 98WO-1100345.  
 XX  
 PR 13-JAN-1998; 98US-0006636.  
 PR 27-JUL-1997; 97IL-0121404.  
 PR 13-JAN-1998; 98US-0006632.  
 XX  
 PA (YISS ) YISSUM RES & DEV CO.  
 XX  
 PI Shani Z, Shoseyov O, Shpiegl E;  
 DR WPI: 1999-180488/15.



AC AAD11046;  
XX  
DT 24-SEP-2001 (first entry)  
DE Chimeric S peptide-cellulose binding domain-S protein encoding DNA.  
XX  
KM Polysaccharide modification; polysaccharide binding domain; PBD; Paper;  
KM yarn; fiber; textile; biological crosslinker; mechanical property;  
KM wet strength; durability; elasticity; cellulose binding domain; CBD;  
KM chimeric protein; S peptide-cellulose binding domain-S protein;  
KM Speg-CBD-Sprot; bovine; ds.  
XX  
OS Chimeric - Clostridium cellulovorans.  
OS Chimeric - Bos sp.  
XX  
PH Key Location/Qualifiers  
FT CDS 1..984  
FT /tag- a  
FT /product= "Speg-CBD-Sprot protein"  
FT /transl\_except= "(pos: 979..981, aa:Xaa)"  
FT /note= "Xaa corresponds to in-frame stop codon;  
FT CDS does not include start and stop codon"  
FT /partial  
FT mISC\_feature /tag- b  
FT 68..624  
FT /note= "this region is derived from C. cellulovorans"  
FT mISC\_feature 652..981  
FT /tag- c  
FT /note= "this region is derived from bovine"  
XX  
PN MO200134091-A2.  
XX  
PD 17-MAY-2001.  
XX  
XX 02-NOV-2000; 2000MO-IL00708.  
XX  
XX 08-NOV-1999; 99US-0164140.  
XX 18-NOV-1999; 99US-0166389.  
XX  
PA (CBDT-) CBD TECHNOLOGIES LTD.  
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
XX  
PI Levy I, Nussinovitch A, Shoseyov O;  
XX  
DR WPI; 2001-457121/49.  
DR P-PSDB; AA05749.  
XX  
XX Preparation of a polysaccharide containing material having at least one  
PT desired structural, chemical, physical, electrical and/or mechanical  
PT property -  
XX  
XX Example 4; Fig 4b-4g; 121pp; English.  
XX  
XX The present invention relates to methods and compositions for cross-  
CC linking and/or modifying the properties of polysaccharide materials.  
CC The method involves treating the polysaccharide structure with a  
CC polysaccharide binding domain (PBD) fusion protein. The method is  
CC used to alter the structural, chemical, physical, electrical and  
CC mechanical properties of polysaccharide materials such as paper,  
CC yarns, fibers and textiles, using biological crosslinking agents.  
CC The polysaccharide containing materials have improved mechanical  
CC properties such as wet strengths, durability and elasticity. The PBD  
CC reagent is applied in the forming stage in fluting paper manufacture  
CC which eliminates the sizing step. The use of a biological crosslinker  
CC improves the recyclability of paper products. The PBD reagent maintains  
CC the fine fibers in a slurry therefore resulting in better recovery of  
CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
CC which enhances the ability of the alkaline glue used in binding to  
CC penetrate paper.  
CC The present sequence is a DNA encoding S peptide-cellulose binding  
CC domain-S protein (Speg-CBD-Sprot), a fusion protein derived from  
CC Clostridium cellulovorans and bovine.  
XX

SQ Sequence 984 BP; 325 A; 240 C; 193 G; 226 T; 0 other;  
Query Match 33.2%; Score 255; DB 22; Length 984;  
Best Local Similarity 100.0%; Pred. No. 7e-123;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 16 TACAACTCTACAAATCAGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75  
DB 118 TACAACTCTACAAATCAGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 177  
QY 76 ACATCTGACAGATTTAAATTTAAATGACGTAAAGTTAGATTATTACACAGTGAT 135  
DB 178 ACATCTGACAGATTTAAATTTAAATGACGTAAAGTTAGATTATTATTACACAGTGAT 237  
QY 136 GGTACACAGACAAACTTCTGTGTGACCAATGCTGTGCTATTATTAGCAATAGTAT 195  
DB 238 GGTACACAGACAAACTTCTGTGTGACCAATGCTGTGCTATTATTAGCAATAGTAT 297  
QY 196 GTTGATTAACACTGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACAGACCCCACTCA 255  
DB 298 GTTGATTAACACTGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACAGACCCCACTCA 357  
QY 256 ACCTATGATACATAT 270  
DB 358 ACCTATGATACATAT 372  
RESULT 10  
AA227550  
ID AA227550 standard; DNA; 1146 BP.  
XX  
XX AA227550:  
XX  
DT 13-DEC-1999 (first entry)  
XX  
XX Gausia luciferase fusion protein coding sequence.  
DE  
XX  
KW Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;  
KW bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;  
KW body paint; squirt gun; balloon; slimy play material; soap; toothpaste;  
KW fusion protein; ds.  
XX  
XX Gausia sp.  
OS  
OS  
PN WO949019-A2.  
XX  
XX 30-SEP-1999.  
PD  
XX  
XX 26-MAR-1999; 99MO-US06698.  
PF  
XX  
XX 27-MAR-1998; 98US-0079624.  
PR 15-JUN-1998; 98US-0089367.  
PR 01-OCT-1998; 98US-0102939.  
XX  
PA (PROL-) PROLUME LTD.  
PA (BRYA/) BRYAN B J.  
XX  
XX Bryan BJ, Szent-Gyorgyi C;  
XX  
XX  
DR P-PSDB; AAY39952.  
XX  
XX New isolated Renilla mulleri, Gausia and Pleuromamma luciferase and  
PT Renilla and Ptilosarcus green fluorescent protein nucleic acids -  
XX  
XX  
XX Disclosure; Page 222-223; 233pp; English.  
XX  
XX This sequence encodes a luciferase of the invention. The invention  
CC relates to Renilla mulleri, Gausia and Pleuromamma luciferase and  
CC Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and  
CC proteins. The luciferases and GFPs can be used in  
CC bioluminescence-generating systems, assays, screening methods, diagnostic  
CC method and articles of manufacture. They can be expressed using

```
CC e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla
CC milleri, Gaussia and Pleuromamma luciferase or Renilla or Ptilosarcus
CC GFPs can be used in e.g. toys, cosmetics, fountains, personal care items,
CC fairy dust, beverages, soft drinks, foods, textile products, bubbles,
CC balloons, personal items, dentifrices, soaps, body paints, bubble bath,
CC ink or paper products. In particular, they can be used in e.g. squirt
CC guns, pellet guns, finger paints, foot bags, greeting cards, slimy play
CC material, clothing, bubble making toys, bath powders, cosmetics, body
CC lotions, gels, body powders, body creams, toothpastes, mouthwashes,
CC soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,
CC frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes,
CC ice, dry ice or fountains. The nucleic acids can also be used to produce
CC transgenic fish and plants.
XX
SQ Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;

Query Match          33.2%; Score 255; DB 20; Length 1146;
Best Local Similarity 100.0%; Pred. No. 7e-123;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
    |||||||
DB 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
QY 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTATGATTTATACACAGATGAT 135
    |||||||
DB 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTATGATTTATACACAGATGAT 135
QY 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGGAATAGCTAT 195
    |||||||
DB 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGGAATAGCTAT 195
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
    |||||||
DB 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
QY 256 ACCTATGATACATAT 270
    |||||||
DB 256 ACCTATGATACATAT 270

RESULT 11
AAD22201
ID AAD22201 standard; DNA; 1146 BP.
XX
AC AAD22201;
XX
DT 12-FEB-2002 (first entry)
XX
DE Gaussia species CBD-luciferase fusion protein encoding DNA.
XX
KW Green fluorescent protein; GFP; bioluminescence generating system; toy;
KW luciferase; finger paint; slimy play material; fishing lure; sparkler;
KW doll; balloon; personal care item; cosmetic; bath powder; body cream;
KW tooth paste; mouth wash; soap; body paint; ornamental transgenic plant;
KW bubble bath; fountain; fairy dust; food; gelatin; icing; frosting; beer;
KW wine; champagne; milk; soft drink; ice cube; ice; foot bag; beverage;
KW dry ice; textile; clothing; paper product; greeting card; wrapping paper;
KW fusion protein; ds.
XX
OS Gaussia sp.
XX
FH Key 1. 1146 Location/Qualifiers
FT CDS
FT /tag= a
FT /product= "Gaussia CBD-luciferase fusion protein"
FT /note= "CDS does not include start and stop codon"
FT /partial
XX
PN WO200168824-A2.
XX
PD 20-SEP-2001.
XX
```

```
PF 15-MAR-2001; 2001WO-US08277.
XX
XX 15-MAR-2000; 2000US-189691P.
XX
XX (PROL-) PROLUME LTD.
XX (BRYA-) BRYAN B J.
XX
PI Bryan BJ, Szent-Gyorgyi C, Szczepaniak W;
XX
XX WPI, 2002-010561/01;
XX P-PSDB; AAE1383.
XX
PT Nucleic acids encoding Renilla reniformis green fluorescent proteins,
XX useful in diagnostic bioluminescence procedures -
XX
PS Disclosure; Page 162-163; 175pp; English.
XX
XX The patent discloses sea pansy (Renilla reniformis) green fluorescent
XX proteins (GFP) and their corresponding polynucleotides. The invention
XX also relates to sequences of the bioluminescence generating system
XX (e.g. luciferase). R. reniformis GFP are used in diagnostic methods
XX and in the production of novelty items such as toys (e.g. squirt gun,
XX pellet gun, toy "Halloween" eggs, toy cigarettes, board/card game
XX toys), finger paints, slimy play material, bubbles in bubble making
XX CC toys, fishing lures, dolls, sparklers, magic wand toys, balloons,
XX CC personal care item (e.g. cosmetic, bath powders, body creams, tooth
XX CC pastes, mouth wash, soaps, body paints, bubble bath), ornamental
XX CC transgenic plants, fountain, fairy dust, food (gelatins, icings,
XX CC frosting, beers, wine, champagne, milk, soft drinks, ice cubes, ice,
XX CC dry ice, beverage), textile (foot bag, clothing) and/or paper product
XX CC (greeting cards, wrapping paper). The present sequence is a DNA
XX encoding Gaussia species CBD-luciferase fusion protein.
XX
SQ Sequence 1146 BP; 396 A; 222 C; 247 G; 281 T; 0 other;

Query Match          33.2%; Score 255; DB 24; Length 1146;
Best Local Similarity 100.0%; Pred. No. 7e-123;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
    |||||||
DB 16 TACAACCTCTAACAAATGACGACAAACAACTCAATTACACCAATTAATCAAAATTACTAAC 75
QY 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTATGATTTATACACAGATGAT 135
    |||||||
DB 76 ACATCTGACAGTGAATTAATTAATGAGCTAAAGTATGATTTATACACAGATGAT 135
QY 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGGAATAGCTAT 195
    |||||||
DB 136 GGTACACAGAGCAAACTTTCTGCTGTGACCATCTGCTGTCATTATTAGGAATAGCTAT 195
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
    |||||||
DB 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAAGCAAGCCCAACATCA 255
QY 256 ACCTATGATACATAT 270
    |||||||
DB 256 ACCTATGATACATAT 270

RESULT 12
AAD11045
ID AAD11045 standard; DNA; 1288 BP.
XX
AC AAD11045;
XX
DT 24-SEP-2001 (first entry)
XX
DE Clostridium cellulovorans protein A-cellulose binding domain DNA.
XX
XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
KW yarn; fiber; textile; biological crosslinker; mechanical property;
KW wet strength; durability; elasticity; protein A-cellulose binding domain;
XX
```

Query Match	15.4%	Score 118:	DB 22:	Length 118:
Best Local Similarity	100.0%	Pred. NO. 2.4e-51.		
Matches 118: Conservative	0:	Mismatches	0:	Indels
			0:	Gaps
				0:

QY 493 GGTGAGACATGACGACGAGTAATCAACTGACCACTGCTCTGGAAAAAGTGAA 552  
 |||||||  
 DB 118 GGTGAGACATGACGACGAGTAATCAACTGACCACTGCTCTGGAAAAAGTGAA 59  
 QY 553 GCGGCGAGGTACTGCTGTCGATCCGACACCGGTAACTACTCTGTGAAC 610  
 |||||||  
 DB 58 GCGGCGAGGTACTGCTGTCGATCCGACACCGGTAACTACTCTGTGAAC 1

## RESULT 14

AAF86252  
 ID AAF86252 standard; DNA: 110 BP.

AC AAF86252;

DT 11-JUL-2001 (first entry)

DE PCR primer #2 used in cloning an optimisation of Ospa gene.

XX Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;  
 SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PE 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PS (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZV/) KUZV M A.

PI Kay WW, Burian J, Kuzv MA;

DR WPI: 2001-316844/34.

PT Method for protecting polkiothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis

PS Example 3; Fig 4B; 35pp; English.

XX This invention relates to a method for the protection against infection  
 CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic fragment of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkiothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents a PCR primer used in the cloning and optimisation of  
 CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the  
 CC invention.

SQ Sequence 110 BP; 13 A; 28 C; 45 G; 24 T; 0 other;

Query Match 13.5%; Score 104; DB 22; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-44;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 TGGTGGCGCCGAGAACTTACGCCGAGAGTGGCCGCGACACCGGTGGGTTGG 397  
 |||||||

DB 1 TGGGTTGGCGCCGAGAACTTACGCCGAGAGTGGCCGCGACACCGGTGGGTTGG 60  
 |||||||

QY 398 GCGGTGTGGCGGCGAGCTTGTGTAAGAGCTCTGTGCTGTG 441  
 |||||||

DB 61 GCGGTGTGGCGGCGAGCTTGTGTAAGAGCTCTGTGCTGTG 104  
 |||||||

RESULT 15  
 AAF86252/C  
 ID AAF86252 standard; DNA: 102 BP.

AC AAF86252;

DT 11-JUL-2001 (first entry)

DE PCR primer #5 used in cloning an optimisation of Ospa gene.

XX Polkiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;  
 SRS; 17E2; fusion construct; PCR primer; ss.

OS Piscirickettsia salmonis.

PN CA2281913-A1.

PD 17-MAR-2001.

PE 17-SEP-1999; 99CA-2281913.

PR 17-SEP-1999; 99CA-2281913.

PS (KAYW/) KAY W W.

PA (BURI/) BURIAN J.

PA (KUZV/) KUZV M A.

PI Kay WW, Burian J, Kuzv MA;

DR WPI: 2001-316844/34.

PT Method for protecting polkiothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis

PS Example 3; Fig 4B; 35pp; English.

XX This invention relates to a method for the protection against infection  
 CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkiothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents a PCR primer used in the cloning and optimisation of  
 CC the P. salmonis Ospa gene. The Ospa gene is used in the method of the  
 CC invention.

SQ Sequence 102 BP; 14 A; 23 C; 30 G; 35 T; 0 other;

Query Match 13.3%; Score 102; DB 22; Length 102;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-43;  
 Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 ACCTACGAGCGTTACAAACAAGAGAGCGGTACGACGATCTGCCGATTTACGAC 678  
 |||||||

DB 102 ACCTACGAGCGTTACAAACAAGAGAGCGGTACGACGATCTGCCGATTTACGAC 43  
 |||||||

QY 679 AAGGCATGATGCGAGGTCAAGAAACAGGAATCTACGGCAC 720  
 |||||||

DB 42 AAGGCATGATGCGAGGTCAAGAAACAGGAATCTACGGCAC 1  
 |||||||

Search completed: October 27, 2002, 20:42:33  
 Job time : 197.542 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: October 27, 2002, 18:35:09 : Search time 1486.04 Seconds  
(without alignments)  
10815.053 Million cell updates/sec

Title: US-09-677-374-5

Perfect score: 768  
Sequence: 1 atgtcagttggaattctacaa.....aggltgattagaccgaaaaa 768Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

1	255	33.2	486	6	AR001082	AR001082 Sequence
2	255	33.2	486	6	AR001083	AR001083 Sequence
3	255	33.2	486	6	AR027448	AR027448 Sequence
4	255	33.2	486	6	AR027449	AR027449 Sequence
5	255	33.2	486	6	AR058257	AR058257 Sequence
6	255	33.2	486	6	AR058258	AR058258 Sequence
7	255	33.2	486	6	118514	118514 Sequence 1
8	255	33.2	486	6	118515	118515 Sequence 3
9	255	33.2	486	6	166614	166614 Sequence 1
10	255	33.2	486	6	166615	166615 Sequence 3
11	255	33.2	486	6	188789	188789 Sequence 1
12	255	33.2	486	6	188790	188790 Sequence 3
13	255	33.2	486	6	AR096204	AR096204 Sequence
14	255	33.2	1146	6	AR151733	AR151733 Sequence
15	255	33.2	1146	6	AX250579	AX250579 Sequence
16	255	33.2	6760	1	CLOCBP	M73817 C10stridium
17	23	3.0	243340	2	AL663088	AL663088 MUS muscu
18	22	2.9	589	8	HJRNA1956	AJ001956 HYlomecon
19	22	2.9	594	8	CMRNA1959	AJ001959 Chelidoni
20	22	2.9	596	8	CMRNA1958	AJ001958 Chelidoni
21	22	2.9	596	8	CMRNA1960	AJ001960 Chelidoni
22	22	2.9	648	8	AF169233	AF169233 Zizania t
23	22	2.9	7286	1	AE005332	AE005332 Escherich
24	22	2.9	10029	1	AE005325	AE005325 Escherich
25	22	2.9	11809	1	AE005290	AE005290 Escherich
26	22	2.9	13505	1	AE005441	AE005441 Escherich
27	22	2.9	13624	1	AE006459	AE006459 Escherich
28	22	2.9	103608	9	AL139161	AL139161 Human DNA
29	22	2.9	222605	1	AP002555	AP002555 Escherich
30	22	2.9	286485	1	AP002557	AP002557 Escherich
31	22	2.9	291136	1	AP002556	AP002556 Escherich
32	21	2.7	702	8	AGHGBEHA	L13954 Agoseris he
33	21	2.7	705	8	MITGBEHA	L13955 Microseris
34	21	2.7	706	8	KIRGBEHA	L20464 Krigia bifl
35	21	2.7	706	8	KIRGBEHA	L13948 Krigia bifl
36	21	2.7	706	8	LAURGBEHA	L13957 Lactuca sat
37	21	2.7	708	8	AF422126	AF422126 Microseri
38	21	2.7	748	8	MCU69708	U69708 Microseris
39	21	2.7	750	8	AF386496	AF386496 Microseri
40	21	2.7	750	8	AF386497	AF386497 Microseri
41	21	2.7	750	8	AF386498	AF386498 Microseri
42	21	2.7	750	8	AF386499	AF386499 Microseri
43	21	2.7	750	8	AF386500	AF386500 Microseri
44	21	2.7	750	8	AF386501	AF386501 Microseri
45	21	2.7	750	8	AF386502	AF386502 Microseri

## ALIGNMENTS

RESULT 1  
AR001082  
LOCUS AR001082  
DEFINITION Sequence 1 from patent US 5738984.  
ACCESSION AR001082  
VERSION AR001082.1 GI:3963149  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov O.  
TITLE Kits and methods of detection using cellulose binding domain fusion  
JOURNAL Patent: US 5738984-A 1 14-APR-1998;  
FEATURES Location/Qualifiers  
source 1..486  
/organism="unknown"

BASE COUNT	ORIGIN
194 a	89 c 79 g 124 t

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;

Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 75  
|||||  
31 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 90

76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 135  
|||||  
91 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 150

136 GGTACACAGGACAACTTCTGTTGACCACTGCTGTCATTTATAGAAATAGCTAT 195  
|||||  
151 GGTACACAGGACAACTTCTGTTGACCACTGCTGTCATTTATAGAAATAGCTAT 210

196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACGACAGCCCAATCA 255  
|||||  
211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACGACAGCCCAATCA 270

256 ACCTATGATACATAT 270  
|||||  
271 ACCTATGATACATAT 285

RESULT 2  
AR001083/c AR001083 486 bp DNA linear PAT 04-DEC-1998  
LOCUS Sequence 3 from patent US 5738984.  
DEFINITION AR001083  
ACCESSION AR001083  
VERSION AR001083.1 GI:3963150  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 486)  
Shoseyov, O.  
TITLE Kits and methods of detection using cellulose binding domain fusion proteins  
JOURNAL Patent: US 5738984-A 3 14-APR-1998;  
FEATURES  
source Location/Qualifiers  
1..486 /organism="unknown"  
BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139; Indels 0; Gaps 0;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 75  
|||||  
456 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 397

76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 135  
|||||  
396 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 337

136 GGTACACAGGACAACTTCTGTTGACCACTGCTGTCATTTATAGAAATAGCTAT 195  
|||||  
336 GGTACACAGGACAACTTCTGTTGACCACTGCTGTCATTTATAGAAATAGCTAT 277

196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACGACAGCCCAATCA 255  
|||||  
276 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACGACAGCCCAATCA 217

256 ACCTATGATACATAT 270  
|||||  
216 ACCTATGATACATAT 202

RESULT 3  
AR027448 AR027448 486 bp DNA linear PAT 29-SEP-1999  
LOCUS Sequence 1 from patent US 5856201.  
DEFINITION

ACCESSION AR027448  
VERSION AR027448.1 GI:5938268  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 486)  
Shoseyov, O., Shpiegel, I., Goldstein, M.A. and Dol, R.H.  
TITLE Methods of detection using a cellulose binding domain fusion product  
JOURNAL Patent: US 5856201-A 1 05-JAN-1999;  
FEATURES  
source Location/Qualifiers  
1..486 /organism="unknown"  
BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139; Indels 0; Gaps 0;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 75  
|||||  
31 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 90

76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 135  
|||||  
91 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 150

136 GGTACACAGGACAACTTCTGTTGACCACTGCTGTCATTTATAGAAATAGCTAT 195  
|||||  
151 GGTACACAGGACAACTTCTGTTGACCACTGCTGTCATTTATAGAAATAGCTAT 210

196 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACGACAGCCCAATCA 255  
|||||  
211 GTTGATTAACACTAGCAAAAGTGACAGCAAACTTCTGTTAAAGAAACGACAGCCCAATCA 270

256 ACCTATGATACATAT 270  
|||||  
271 ACCTATGATACATAT 285

RESULT 4  
AR027449/c AR027449 486 bp DNA linear PAT 29-SEP-1999  
LOCUS Sequence 3 from patent US 5856201.  
DEFINITION AR027449  
ACCESSION AR027449  
VERSION AR027449.1 GI:5938269  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 486)  
Shoseyov, O., Shpiegel, I., Goldstein, M.A. and Dol, R.H.  
TITLE Methods of detection using a cellulose binding domain fusion product  
JOURNAL Patent: US 5856201-A 3 05-JAN-1999;  
FEATURES  
source Location/Qualifiers  
1..486 /organism="unknown"  
BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139; Indels 0; Gaps 0;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 75  
|||||  
456 TACAACTCTAACAAATCAGCACAACAACTCAATTACACCAATATCAAAATTTACTAAC 397

76 ACATCTGACAGTGATTTAAATTTAAATGACGTAAGTAAGTATTTATTTACACAGTGAT 135  
|||||



Db 396 ACATCTGCAGATGATTTAAATTAATGACGTAAAGCTAGATATATATACACAAAGTAT 337  
QY 136 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 195  
Db 336 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 277  
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255  
Db 276 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 217  
QY 256 ACCTATGATACATAT 270  
Db 216 ACCTATGATACATAT 202

RESULT 5  
AR058257  
LOCUS AR058257 486 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 1 from patent US 5837814.  
ACCESSION AR058257  
VERSION AR058257.1 GI:5983834  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.  
TITLE Cellulose binding domain proteins  
JOURNAL Patent: US 5837814-A 1 17-NOV-1998;  
FEATURES  
source 1..486  
BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 75  
Db 31 TACAACCTTAACAATGACACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 90  
QY 76 ACATCTGCAGATGATTTAAATTAATGAGCTAAAGTATTAATTAACAAAGTAT 135  
Db 91 ACATCTGCAGATGATTTAAATTAATGAGCTAAAGTATTAATTAACAAAGTAT 150  
QY 136 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 195  
Db 151 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 210  
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255  
Db 211 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 270  
QY 256 ACCTATGATACATAT 270  
Db 271 ACCTATGATACATAT 285

RESULT 6  
AR058258/c  
LOCUS AR058258 486 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5837814.  
ACCESSION AR058258  
VERSION AR058258.1 GI:5983835  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M. and Dol,R.  
TITLE Cellulose binding domain proteins

JOURNAL Patent: US 5837814-A 3 17-NOV-1998;  
FEATURES  
source 1..486  
BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 75  
Db 456 TACAACCTTAACAATGACACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 397  
QY 76 ACATCTGCAGATGATTTAAATTAATGAGCTAAAGTATTAATTAACAAAGTAT 135  
Db 396 ACATCTGCAGATGATTTAAATTAATGAGCTAAAGTATTAATTAACAAAGTAT 337  
QY 136 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 195  
Db 336 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 277  
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255  
Db 276 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 217  
QY 256 ACCTATGATACATAT 270  
Db 216 ACCTATGATACATAT 202

RESULT 7  
I18514  
LOCUS I18514 486 bp DNA linear PAT 07-OCT-1996  
DEFINITION Sequence 1 from patent US 5496934.  
ACCESSION I18514  
VERSION I18514.1 GI:1598869  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Dol,R.H.  
TITLE Nucleic acids encoding a cellulose binding domain  
JOURNAL Patent: US 5496934-A 1 05-MAR-1996;  
FEATURES  
source 1..486  
BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGACACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 75  
Db 31 TACAACCTTAACAATGACACAAACAACTCAATTAACCAATATCAAAATTAAGTAC 90  
QY 76 ACATCTGCAGATGATTTAAATTAATGAGCTAAAGTATTAATTAACAAAGTAT 135  
Db 91 ACATCTGCAGATGATTTAAATTAATGAGCTAAAGTATTAATTAACAAAGTAT 150  
QY 136 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 195  
Db 151 GGTACACAAAGACAACTTCTGCTGACCACTGCTGATTAATAGAAATAGCTAT 210  
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 255  
Db 211 GTTGATTAACACTAGCAAGTGACAGCAAACTTCTGTTAAAGAAACAGCAAGCCCAACATCA 270  
QY 256 ACCTATGATACATAT 270

Db 271 ACCTATGATACATAT 285

RESULT 8  
LOCUS 118515/c  
DEFINITION Sequence 3 from patent US 5496934.  
ACCESSION 118515  
VERSION 118515.1 GI:1598870  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shipiegl,I., Goldstein,M.A. and Dol,R.H.  
TITLE Nucleic acids encoding a cellulose binding domain  
JOURNAL Patent: US 5496934-A 3 05-MAR-1996;  
FEATURES Location/Qualifiers  
source 1..486

BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACACTCTAACAATGACGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 75  
|||||  
Db 456 TACAACTCTAACAATGACGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 397  
|||||

QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGTAAGTATTTATTTACACAAGTGAT 135  
|||||  
Db 396 ACATCTGACAGTGATTTAAATTTAAATGAGTAAGTATTTATTTACACAAGTGAT 337  
|||||

QY 136 GGTACACAGAGCAAACTTCTGCTGTGACCATCTGCTGATTTATTTAGAAATAGCTAT 195  
|||||  
Db 336 GGTACACAGAGCAAACTTCTGCTGTGACCATCTGCTGATTTATTTAGAAATAGCTAT 277  
|||||

QY 196 GTTGATTAACACTAGCAAGGACGCAAACTGCTTTAAAGAAACGAGCCCAACATCA 255  
|||||  
Db 276 GTTGATTAACACTAGCAAGGACGCAAACTGCTTTAAAGAAACGAGCCCAACATCA 217  
|||||

QY 256 ACCTATGATACATAT 270  
|||||  
Db 216 ACCTATGATACATAT 202  
|||||

RESULT 9  
LOCUS 166614  
DEFINITION Sequence 1 from patent US 5670623.  
ACCESSION 166614  
VERSION 166614.1 GI:2724592  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shipiegl,I., Goldstein,M.A. and Dol,R.H.  
TITLE Methods of use of cellulose binding domain proteins  
JOURNAL Patent: US 5670623-A 1 23-SEP-1997;  
FEATURES Location/Qualifiers  
source 1..486

BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACTCTAACAATGACGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 75  
|||||  
Db 31 TACAACTCTAACAATGACGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 90  
|||||

QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGTAAGTATTTATTTACACAAGTGAT 135  
|||||  
Db 91 ACATCTGACAGTGATTTAAATTTAAATGAGTAAGTATTTATTTACACAAGTGAT 150  
|||||

QY 136 GGTACACAGAGCAAACTTCTGCTGTGACCATCTGCTGATTTATTTAGAAATAGCTAT 195  
|||||  
Db 151 GGTACACAGAGCAAACTTCTGCTGTGACCATCTGCTGATTTATTTAGAAATAGCTAT 210  
|||||

QY 196 GTTGATTAACACTAGCAAGGACGCAAACTGCTTTAAAGAAACGAGCCCAACATCA 255  
|||||  
Db 211 GTTGATTAACACTAGCAAGGACGCAAACTGCTTTAAAGAAACGAGCCCAACATCA 270  
|||||

QY 256 ACCTATGATACATAT 270  
|||||  
Db 271 ACCTATGATACATAT 285  
|||||

RESULT 10  
LOCUS 166615/c  
DEFINITION Sequence 3 from patent US 5670623.  
ACCESSION 166615  
VERSION 166615.1 GI:2724593  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shipiegl,I., Goldstein,M.A. and Dol,R.H.  
TITLE Methods of use of cellulose binding domain proteins  
JOURNAL Patent: US 5670623-A 3 23-SEP-1997;  
FEATURES Location/Qualifiers  
source 1..486

BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACTCTAACAATGACGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 75  
|||||  
Db 456 TACAACTCTAACAATGACGCAACAACAACTCAATTACCAATTAATCAAAATTACTAAC 397  
|||||

QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGTAAGTATTTATTTACACAAGTGAT 135  
|||||  
Db 396 ACATCTGACAGTGATTTAAATTTAAATGAGTAAGTATTTATTTACACAAGTGAT 337  
|||||

QY 136 GGTACACAGAGCAAACTTCTGCTGTGACCATCTGCTGATTTATTTAGAAATAGCTAT 195  
|||||  
Db 336 GGTACACAGAGCAAACTTCTGCTGTGACCATCTGCTGATTTATTTAGAAATAGCTAT 277  
|||||

QY 196 GTTGATTAACACTAGCAAGGACGCAAACTGCTTTAAAGAAACGAGCCCAACATCA 255  
|||||  
Db 276 GTTGATTAACACTAGCAAGGACGCAAACTGCTTTAAAGAAACGAGCCCAACATCA 217  
|||||

QY 256 ACCTATGATACATAT 270  
|||||  
Db 216 ACCTATGATACATAT 202  
|||||

RESULT 11  
LOCUS 188789  
DEFINITION Sequence 1 from patent US 5719044.  
ACCESSION 188789  
VERSION 188789.1 GI:3408729  
KEYWORDS

BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Dol,R.H.  
TITLE Cellulose binding domain fusion proteins  
JOURNAL Patent: US 5719044-A 1 17-FEB-1998;  
FEATURES Location/Qualifiers  
source 1..486  
BASE COUNT 194 a 89 c 79 g 124 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGAGCACAACAACATTCATTAACCAATATCAAAATTAAGTACTAAC 75  
DB 31 TACAACCTTAACAATGAGCACAACAACATTCATTAACCAATATCAAAATTAAGTACTAAC 90  
QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135  
DB 91 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 150  
QY 136 GGTACACAGGACAACTTCTGTGTGACCATGCTGTGCATATTATAGAAATAGCTAT 195  
DB 151 GGTACACAGGACAACTTCTGTGTGACCATGCTGTGCATATTATAGAAATAGCTAT 210  
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAACAGCAAGCCCAACATCA 255  
DB 211 GTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAACAGCAAGCCCAACATCA 270  
QY 256 ACCTATGATACATAT 270  
DB 271 ACCTATGATACATAT 285

RESULT 12  
188790/c 188790 486 bp DNA linear PAT 10-AUG-1998  
LOCUS Sequence 3 from patent US 5719044.  
DEFINITION 188790  
ACCESSION 188790.1 GI:3408730  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Shoseyov,O., Shpiegl,I., Goldstein,M.A. and Dol,R.H.  
TITLE Cellulose binding domain fusion proteins  
JOURNAL Patent: US 5719044-A 3 17-FEB-1998;  
FEATURES Location/Qualifiers  
source 1..486  
BASE COUNT 124 a 79 c 89 g 194 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 486;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGAGCACAACAACATTCATTAACCAATATCAAAATTAAGTACTAAC 75  
DB 456 TACAACCTTAACAATGAGCACAACAACAATTCATTAACCAATATCAAAATTAAGTACTAAC 397  
QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135  
DB 396 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 337  
QY 136 GGTACACAGGACAACTTCTGTGTGACCATGCTGTGCATATTATAGAAATAGCTAT 195  
DB 336 GGTACACAGGACAACTTCTGTGTGACCATGCTGTGCATATTATAGAAATAGCTAT 277

QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAACAGCAAGCCCAACATCA 255  
DB 276 GTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAACAGCAAGCCCAACATCA 217  
QY 256 ACCTATGATACATAT 270  
DB 216 ACCTATGATACATAT 202

RESULT 13  
AR096204 499 bp DNA linear PAT 08-SEP-2000  
LOCUS Sequence 5 from patent US 6005092.  
DEFINITION AR096204  
ACCESSION AR096204  
VERSION AR096204.1 GI:10024795  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 499)  
AUTHORS Shoseyov,O. and Shanl,Z.  
TITLE Arabidopsis thaliana endo-1,4-.beta.-glucanase gene and promoter  
JOURNAL Patent: US 6005092-A 5 21-DEC-1999;  
FEATURES Location/Qualifiers  
source 1..499  
BASE COUNT 197 a 93 c 82 g 127 t  
ORIGIN

Query Match 33.2%; Score 255; DB 6; Length 499;  
Best Local Similarity 100.0%; Pred. No. 2.3e-139;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TACAACCTTAACAATGAGCACAACAACATTCATTAACCAATATCAAAATTAAGTACTAAC 75  
DB 36 TACAACCTTAACAATGAGCACAACAACATTCATTAACCAATATCAAAATTAAGTACTAAC 95  
QY 76 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 135  
DB 96 ACATCTGACAGTGATTTAAATTTAAATGAGCTAAAGTTAGATATTATTACACAAGTGAT 155  
QY 136 GGTACACAGGACAACTTCTGTGTGACCATGCTGTGCATATTATAGAAATAGCTAT 195  
DB 156 GGTACACAGGACAACTTCTGTGTGACCATGCTGTGCATATTATAGAAATAGCTAT 215  
QY 196 GTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAACAGCAAGCCCAACATCA 255  
DB 216 GTTGATTAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAACAGCAAGCCCAACATCA 275  
QY 256 ACCTATGATACATAT 270  
DB 276 ACCTATGATACATAT 290

RESULT 14  
AR151733 1146 bp DNA linear PAT 08-AUG-2001  
LOCUS Sequence 21 from patent US 6232107.  
DEFINITION AR151733  
ACCESSION AR151733  
VERSION AR151733.1 GI:15117783  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1146)  
AUTHORS Bryan,B.J. and Szent-Gyorgyi,C.  
TITLE Luciferases, fluorescent proteins, nucleic acids encoding the  
diagnostics, high throughput screening and the use thereof in  
Journal Patent: US 6232107-A 21 15-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..1146

16 TACAACTCTAACAACATCAGCACAACAACTCAATTACACCAATATTCAAATTTACTAAC 75

Search completed: October 27, 2002, 21:53:00  
Job time : 1490.04 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 : Search time 13.2642 Seconds  
(without alignments)  
471.414 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335  
Sequence: 1 MSVEFYNSKSAQNTSTPTI.....IYGTACPOPDGRMOVISTEK 256

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5a\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5b\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6a\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6b\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCOTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480.5	36.0	382	4	US-09-277-716-22 Sequence 22, Appl
2	477	35.7	162	1	US-08-048-164A-2 Sequence 2, Appl
3	477	35.7	162	1	US-08-460-462-2 Sequence 2, Appl
4	477	35.7	162	1	US-08-460-457-2 Sequence 2, Appl
5	477	35.7	162	1	US-08-460-458-2 Sequence 2, Appl
6	477	35.7	162	2	US-08-460-455-2 Sequence 2, Appl
7	477	35.7	162	2	US-08-330-394A-2 Sequence 2, Appl
8	477	35.7	163	3	US-09-006-636-7 Sequence 7, Appl
9	477	35.7	163	4	US-09-006-632-7 Sequence 7, Appl
10	472	35.4	154	2	US-08-330-394A-29 Sequence 29, Appl
11	472	35.4	156	2	US-08-330-394A-22 Sequence 22, Appl
12	245	18.4	167	5	PCT-US95-13813-9 Sequence 9, Appl
13	245	18.4	493	4	US-09-198-956-10 Sequence 10, Appl
14	245	18.4	493	4	US-09-198-955A-12 Sequence 12, Appl
15	149.5	11.2	1436	4	US-09-136-574A-43 Sequence 43, Appl
16	148.5	11.1	616	4	US-09-136-574A-47 Sequence 47, Appl
17	148.5	11.1	1751	4	US-09-136-574A-44 Sequence 44, Appl
18	122.5	9.2	551	2	US-09-033-573A-1 Sequence 1, Appl
19	100.5	7.5	700	2	US-07-862-588A-2 Sequence 2, Appl
20	85	6.4	1021	1	US-07-910-760-12 Sequence 12, Appl
21	85	6.4	1021	4	US-08-440-519-12 Sequence 12, Appl
22	85	6.4	1021	4	US-08-440-519-12 Sequence 12, Appl
23	83	6.2	382	4	US-08-444-818-68 Sequence 68, Appl
24	83	6.2	460	4	US-08-444-818-20 Sequence 20, Appl
25	83	6.2	512	4	US-08-867-611-58 Sequence 58, Appl
26	83	6.2	592	4	US-08-867-611-47 Sequence 47, Appl
27	83	6.2	594	4	US-08-867-611-48 Sequence 48, Appl

28	83	6.2	597	4	US-08-867-611-16 Sequence 16, Appl
29	83	6.2	597	5	PCT-US92-06965A-21 Sequence 21, Appl
30	83	6.2	599	4	US-08-867-611-18 Sequence 18, Appl
31	83	6.2	599	5	PCT-US92-06965A-23 Sequence 23, Appl
32	83	6.2	739	4	US-08-867-611-49 Sequence 49, Appl
33	83	6.2	739	4	US-08-444-818-148 Sequence 148, App
34	83	6.2	781	4	US-08-867-611-4 Sequence 4, Appl
35	83	6.2	781	5	PCT-US92-06965A-9 Sequence 9, Appl
36	83	6.2	859	4	US-08-444-818-30 Sequence 30, Appl
37	83	6.2	970	1	US-08-375-709-7 Sequence 7, Appl
38	83	6.2	970	1	US-08-752-929-7 Sequence 7, Appl
39	83	6.2	970	4	US-09-090-793-5 Sequence 5, Appl
40	83	6.2	971	4	US-08-867-611-52 Sequence 52, Appl
41	83	6.2	971	4	US-08-867-611-53 Sequence 53, Appl
42	83	6.2	992	4	US-08-867-611-54 Sequence 54, Appl
43	83	6.2	1786	4	US-08-444-818-54 Sequence 54, Appl
44	83	6.2	2261	4	US-08-444-818-66 Sequence 66, Appl
45	83	6.2	2772	4	US-08-444-818-89 Sequence 89, Appl

ALIGNMENTS

```
RESULT 1
US-09-277-716-22
; Sequence 22, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Artificial Sequence: fusion protein
; FEATURE:
; OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia luciferase fusion prote
US-09-277-716-22
Query Match 36.0%; Score 480.5; DB 4; Length 382;
Best Local Similarity 62.6%; Pred. No. 4.1e-41;
Matches 107; Conservative 10; Mismatches 33; Indels 21; Gaps 4;
OY 1 MSVEFYNSKSAQNTSTPTIIRKINTSDSDLNLDVVRVRYTSDGQGGCFWCDHMGAL 60
Db 1 MSVEFYNSKSAQNTSTPTIIRKINTSDSDLNLDVVRVRYTSDGQGGCFWCDHMGAL 60
OY 61 LGNSYVNTSKVTANFKEFASPTSTYDYLDPSHMRCL---CGSSLITISFLVGCAG 117
Db 61 LGNSYVNTSKVTANFKEFASPTSTYDYLDPSHMRCL---CGSSLITISFLVGCAG 117
OY 118 NESRQ----EVGATGAVGCVAGQLFGKSGRVSNAITGA-VLGLIGS 162
Db 121 NYTOTNDYSFSSSTSTVNVKPYGVY-----IGGAKVLTGAPGS 159
RESULT 2
US-08-048-164A-2
; Sequence 2, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
```



TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 162 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-460-457-2

Query Match 35.7%; Score 477; DB 1; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSVEFYNSKSAQTNSTPIIKITNTSDSLNLDVKVRYRYTSDGTGGQTFWCDHAGAL 60  
DB 6 MSVEFYNSKSAQTNSTPIIKITNTSDSLNLDVKVRYRYTSDGTGGQTFWCDHAGAL 65

OY 61 LGNSYVDNTSKVTANFYKETASPTSTYDTYLD 92  
DB 66 LGNSYVDNTSKVTANFYKETASPTSTYDTYVE 97

RESULT 5  
US-08-460-458-2

; Sequence 2, Application US/08460458  
; Patent No. 5738984

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,458

FILING DATE: concurrently herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/048,164

FILING DATE: 14-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7809-007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 162 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-460-458-2

Query Match 35.7%; Score 477; DB 1; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSVEFYNSKSAQTNSTPIIKITNTSDSLNLDVKVRYRYTSDGTGGQTFWCDHAGAL 60  
DB 6 MSVEFYNSKSAQTNSTPIIKITNTSDSLNLDVKVRYRYTSDGTGGQTFWCDHAGAL 65

OY 61 LGNSYVDNTSKVTANFYKETASPTSTYDTYLD 92  
DB 66 LGNSYVDNTSKVTANFYKETASPTSTYDTYVE 97

RESULT 6  
US-08-460-455-2

; Sequence 2, Application US/08460455

; Patent No. 5837814

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Shoplegl, Itai

APPLICANT: Goldstein, Marc A.

TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,455

FILING DATE: concurrently herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/048,164

FILING DATE: 14-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7809-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 162 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-460-455-2

Query Match 35.7%; Score 477; DB 2; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSVEFYNSKSAQTNSTPIIKITNTSDSLNLDVKVRYRYTSDGTGGQTFWCDHAGAL 60  
DB 6 MSVEFYNSKSAQTNSTPIIKITNTSDSLNLDVKVRYRYTSDGTGGQTFWCDHAGAL 65

OY 61 LGNSYVDNTSKVTANFYKETASPTSTYDTYLD 92  
DB 66 LGNSYVDNTSKVTANFYKETASPTSTYDTYVE 97

RESULT 7  
US-08-330-394A-2

; Sequence 2, Application US/08330394A

; Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Yosef, Karney

APPLICANT: Shoplegl, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: DO1, ROY H.  
TITLE OF INVENTION: METHODS OF DETECTION USING THE  
CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/330,394A  
FILING DATE: 27-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CIP OF PCT/US94/04132  
FILING DATE: 14-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MISROCK, S. LESLIE  
REFERENCE/DOCKET NUMBER: 7809-005  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66441 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-330-394A-2

Query Match 35.7%; Score 477; DB 2; length 162;  
Best Local Similarity 97.8%; Pred. No. 2,7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTQGTFCMDHAGAL 60  
|||||  
DB 6 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTQGTFCMDHAGAL 65  
|||||

QY 61 LGNSYVDNTSKVTANFEKTAFTSTYDYLD 92  
|||||  
DB 66 LGNSYVDNTSKVTANFEKTAFTSTYDYLD 97  
|||||

RESULT 8  
US-09-006-636-7  
Sequence 7, Application US/09006636  
Patent No. 6005092  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shani, Ziv  
TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-  
GLUCANASE GENE, PROMOTER AND PROTEIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/006,636  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7809-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-006-636-7

Query Match 35.7%; Score 477; DB 3; length 163;  
Best Local Similarity 97.8%; Pred. No. 2,7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTQGTFCMDHAGAL 60  
|||||  
DB 7 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTQGTFCMDHAGAL 66  
|||||

QY 61 LGNSYVDNTSKVTANFEKTAFTSTYDYLD 92  
|||||  
DB 67 LGNSYVDNTSKVTANFEKTAFTSTYDYLD 98  
|||||

RESULT 9  
US-09-006-632-7  
Sequence 7, Application US/09006632  
Patent No. 618440  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shani, Ziv  
TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED  
MORPHOLOGY  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/006,632  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7809-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
STRANDEDNESS: single



TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-006-632-7

Query Match 35.7%; Score 477; DB 4; Length 163;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKRYRYYTSDGTGQGFMCDAAGAL 60  
DB 7 MSVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKRYRYYTSDGTGQGFMCDAAGAL 66  
QY 61 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92  
DB 67 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 98

RESULT 10  
US-08-330-394A-29  
Sequence 29, Application US/08330394A  
Patent No. 5856201  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Yosef, Karmey  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldsteil, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: METHODS OF DETECTION USING THE  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/330,394A  
FILING DATE: 27-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CIP OF PCT/US94/04132  
FILING DATE: 14-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MISROCK, S. LESLIE  
REFERENCE/DOCKET NUMBER: 7809-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66441 PENNIE  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 154 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-330-394A-29

Query Match 35.4%; Score 472; DB 2; Length 154;  
Best Local Similarity 97.8%; Pred. No. 8.2e-41;  
Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKRYRYYTSDGTGQGFMCDAAGAL 61  
DB 1 SVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKRYRYYTSDGTGQGFMCDAAGAL 60  
QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92

DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 11  
US-08-330-394A-22  
Sequence 22, Application US/08330394A  
Patent No. 5856201  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Yosef, Karmey  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldsteil, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: METHODS OF DETECTION USING THE  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/330,394A  
FILING DATE: 27-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CIP OF PCT/US94/04132  
FILING DATE: 14-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MISROCK, S. LESLIE  
REFERENCE/DOCKET NUMBER: 7809-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66441 PENNIE  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-330-394A-22

Query Match 35.4%; Score 472; DB 2; Length 156;  
Best Local Similarity 97.8%; Pred. No. 8.4e-41;  
Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKRYRYYTSDGTGQGFMCDAAGAL 61  
DB 1 SVEFYNSKSAQTNSITPIIKITNTSDSLNLDVKKRYRYYTSDGTGQGFMCDAAGAL 60  
QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92  
DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 12  
PCT-US95-13813-9  
Sequence 9, Application PC/RUS9513813  
GENERAL INFORMATION:  
APPLICANT: Yeda Research and Development Co. Ltd.  
APPLICANT: Ramot University Authority for Applied  
APPLICANT: Research and Industrial Development Ltd.  
APPLICANT: Technion Research and Development Foundation Ltd.  
APPLICANT: Bayer, Edward A.  
APPLICANT: Morag, Ely  
APPLICANT: Wilchek, Meir

```
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-13813-9
;
; Query Match 18.4%; Score 245; DB 5; Length 167;
; Best Local Similarity 51.5%; Pred. No. 1.7e-17;
; Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;
;
; Db
; 1 MSVEFYNSKSAQNTNTPPIIKITNTSDSLNDVAVRYRYTSDGOTGOTFCWCDHAGAL 60
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 9 LKVEFYNSPDSNTNINPQFKVTNTGSSAIDLSKILRLRYTYVGGQDKOTFCWCDHA-AI 67
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 61 LGN--SYVDNTSKVTANFYKETASPTSTYDTYLDPSHMRGCLQ 101
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 68 IGSNGSYNGITSNVKGTFFVKMSSS-TNNADTYLEISFTGCTLE 109
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 13
; US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377 200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
```

```
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
; US-09-198-956-10
;
; Query Match 18.4%; Score 245; DB 4; Length 493;
; Best Local Similarity 51.5%; Pred. No. 8.3e-17;
; Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;
;
; Db
; 1 MSVEFYNSKSAQNTNTPPIIKITNTSDSLNDVAVRYRYTSDGOTGOTFCWCDHAGAL 60
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 335 LKVEFYNSPDSNTNINPQFKVTNTGSSAIDLSKILRLRYTYVGGQDKOTFCWCDHA-AI 393
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 61 LGN--SYVDNTSKVTANFYKETASPTSTYDTYLDPSHMRGCLQ 101
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 394 IGSNGSYNGITSNVKGTFFVKMSSS-TNNADTYLEISFTGCTLE 435
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 14
; US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
; US-09-198-955A-12
;
; Query Match 18.4%; Score 245; DB 4; Length 493;
; Best Local Similarity 51.5%; Pred. No. 8.3e-17;
; Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;
;
; Db
; 1 MSVEFYNSKSAQNTNTPPIIKITNTSDSLNDVAVRYRYTSDGOTGOTFCWCDHAGAL 60
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 335 LKVEFYNSPDSNTNINPQFKVTNTGSSAIDLSKILRLRYTYVGGQDKOTFCWCDHA-AI 393
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 61 LGN--SYVDNTSKVTANFYKETASPTSTYDTYLDPSHMRGCLQ 101
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; 394 IGSNGSYNGITSNVKGTFFVKMSSS-TNNADTYLEISFTGCTLE 435
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 15
; US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
```

```

1 GENERAL INFORMATION:
2
3 APPLICANT: Farrington, Graham K.
4
5 Anderson, Paige
6 Gibbs, Moreland
7 Bergquist, Peter
8 Daniels, Roy
9 Morgan, Hugh W.
10 Williams, Diane P.
11
12 TITLE OF INVENTION: Compositions and Methods for
13 Treating Cellulose Containing Fabrics Using Truncated
14 Cellulase Enzyme Compositions
15
16 NUMBER OF SEQUENCES: 49
17
18 CORRESPONDENCE ADDRESS:
19 ADDRESSEE: Howson and Howson
20 STREET: Spring House Corporate Center, P.O. Box 457
21 CITY: Spring House
22 STATE: PA
23 COUNTRY: USA
24 ZIP: 19477
25
26 COMPUTER READABLE FORM:
27
28 MEDIUM TYPE: Diskette
29 COMPUTER: IBM Compatible
30 OPERATING SYSTEM: DOS
31
32 CURRENT APPLICATION DATA:
33 APPLICATION NUMBER: US/09/136,574A
34 FILING DATE: 19-Aug-1998
35
36 CLASSIFICATION: <unknown>
37
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: US 08/932,571
40 FILING DATE: September 19, 1997
41
42 ATTORNEY/AGENT INFORMATION:
43 NAME: Bak, Mary E.
44 REGISTRATION NUMBER: 31,215
45 REFERENCE/DOCKET NUMBER: 199705001/CIP
46
47 TELECOMMUNICATION INFORMATION:
48 TELEPHONE: 215-540-9200
49 TELEFAX: 215-540-5818
50
51 TELEX: <unknown>
52
53 INFORMATION FOR SEQ ID NO: 43:
54
55 SEQUENCE CHARACTERISTICS:
56 LENGTH: 1426 amino acids
57 TYPE: amino acid
58 STRANDEDNESS: single
59 TOPOLOGY: linear
60
61 MOLECULE TYPE: No. 6294366e
62
63 SEQUENCE DESCRIPTION: SEQ ID NO: 43:
64
65 US-09-136-574A-43
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
95
```

Search completed: October 27, 2002, 11:00:30  
Job time : 15.2642 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 : Search time 34.487 Seconds  
(without alignments)  
824.509 Million cell updates/sec

```
Title: US-09-677-374-6
Perfect score: 1335
Sequence: 1 MSVERYNSNKSQAQTNSIMP.....IYGTACPOPDGRMVOISTEK 256
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

A\_Geneseq\_032802.\*

1:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	135	100.0	256	22	AAB81128	C17E2 Ospa constru
2	832	62.3	161	22	AAB81127	Optimised Ospa pro
3	815	61.0	162	22	AA878025	Piscificketsia sa
4	815	61.0	162	22	AAB81126	Ospa antigen amino
5	487	36.5	190	22	AAE05746	Clostridium cellu
6	487	36.5	328	22	AAE05749	Chimeric S peptide
7	487	36.5	341	22	AAE05747	Clostridium cellu
8	483	36.2	162	15	FAK63634	Cellulose binding
9	483	36.2	163	22	AAE05745	Clostridium cellu
10	483	36.2	428	22	AAE05748	Clostridium cellu
11	480.5	36.0	382	20	AAV39952	Gaussia luciferase

12	477	35.7	162	20	AA90077
13	472	35.4	154	20	AA90081
14	472	35.4	156	20	AA90080
15	245	18.4	167	17	AA95018
16	245	18.4	176	21	AA95420
17	245	18.4	493	20	AA928650
18	245	18.4	493	20	AA945218
19	245	18.4	531	18	AA951538
20	245	18.4	1563	19	AA943108
21	149.5	11.1	1426	20	AA913492
22	148.5	11.1	616	20	AA913493
23	148.5	11.1	1751	20	AA913493
24	127.5	9.6	1352	22	AA663862
25	126.5	9.5	1550	22	AA636363
26	122.5	9.2	551	18	AA918790
27	113.5	8.5	499	14	AA944122
28	112	8.4	20	22	AA981130
29	111	8.3	224	22	AA920105
30	102.5	7.7	223	20	AA934487
31	102.5	7.7	230	20	AA934362
32	102	7.6	782	12	AA915625
33	101.5	7.6	154	11	AA905799
34	100.5	7.5	700	12	AA913227
35	98	7.3	922	20	AA934597
36	98	7.3	922	20	AA988419
37	98	7.3	922	21	AA955548
38	95	7.1	309	22	AA915906
39	95	7.1	2309	22	AA966232
40	92	6.9	285	20	AA944168
41	92	6.9	285	21	AA944244
42	92	6.9	285	22	AA929055
43	91.5	6.9	285	21	AA978280
44	91.5	6.9	285	21	AA973440
45	91.5	6.9	285	22	AA993288

## ALIGNMENTS

RESULT	1
AAB81128	
ID	AAB81128 standard; Protein; 256 AA.
XX	
AC	AAB81128;
XX	
DT	11-JUL-2001 (first entry)
XX	
DE	C17E2 OspA construct with N-terminal fusion partner.
XX	
KW	Polikiothermic fish; Piscirickettsia salmonis; rickettsial pathogen;
KW	vaccine; Osprey; salmonid rickettsial septicaemia; rickettsial disease;
KW	SRS; 17E2; fusion construct.
XX	
OS	Piscirickettsia salmonis.
OS	Synthetic.
XX	
FT	Key
FT	Region
FT	Location/Qualifiers
FT	1..95
FT	/label= Undefined_N-terminal_fusion_partner
FT	96..256
FT	/label= C17E2_OspA
FT	/note= "Product of OspA gene optimised for expression in
FT	Escherichia coli"
XX	
FN	CA2281913-A1.
XX	
PD	17-MAR-2001.
XX	
PF	17-SEP-1999; 99CA-2281913.
XX	
PR	17-SEP-1999; 99CA-2281913.
XX	
PA	(KAYW/) KAY W W.

PA (BURJ/) BURIAN J.  
 PA (KUZU/) KUZUK M.A.  
 XX  
 PI Kay WW, Burian J, Kuzuk MA;  
 XX  
 DR WPI: 2001-316844/34.  
 DR N-PSDB: AAF86248.  
 PT Method for protecting polkiothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the ospa protein of Piscirickettsia salmonis  
 XX  
 PS Example 4; Fig 5; 35pp; English.  
 XX This invention relates to a method for the protection against infection  
 CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkiothermic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents the amino acid sequence of Cl7E2, a P. salmonis Ospa  
 CC construct optimised for expression in Escherichia coli, fused to an  
 CC undefined N-terminal fusion partner. The fusion protein is used in a  
 CC vaccine to create an anti-Ospa antibody response.  
 CC  
 SQ Sequence 256 AA:  
 Query Match 100.0%; Score 1335; DB 22; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-122;  
 Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSVEFYNSKSKSQSTSTPIITKTTSDNLNDVKRYRYTSDGTQGTFCMDHAGAL 60  
 DB 1 MSVEFYNSKSKSQSTSTPIITKTTSDNLNDVKRYRYTSDGTQGTFCMDHAGAL 60  
 QY 61 LGNSYDNTSKYANFVKETASPTSTYDYLDPMSHRCLOGLSSLLITSVFLVCAONFS 120  
 DB 61 LGNSYDNTSKYANFVKETASPTSTYDYLDPMSHRCLOGLSSLLITSVFLVCAONFS 120  
 QY 121 ROEVAATGAVVAGVAGQLFGKSGSRVSMAGVAGLIGSKIGOSMDQDKIKLNSL 180  
 DB 121 ROEVAATGAVVAGVAGQLFGKSGSRVSMAGVAGLIGSKIGOSMDQDKIKLNSL 180  
 QY 181 EKKVAGVTRMRNPDGNSYSVEPVRYTORYNKQERRQOYCFEPOOKAMIAQOKOETIGT 240  
 DB 181 EKKVAGVTRMRNPDGNSYSVEPVRYTORYNKQERRQOYCFEPOOKAMIAQOKOETIGT 240  
 QY 241 ACPOPGRMVOYISTEK 256  
 DB 241 ACPOPGRMVOYISTEK 256  
 RESULT 2  
 ID AAB81127 standard; Protein; 161 AA.  
 AC AAB81127;  
 XX 11-JUL-2001 (first entry)  
 DE Optimised Ospa protein 17E2 amino acid sequence.  
 XX  
 KW Polkiothermic fish; piscirickettsia salmonis; rickettsial pathogen;  
 KW vaccine; Ospa; salmonid rickettsial septicemia; rickettsial disease;  
 KW SRS.  
 XX  
 OS Piscirickettsia salmonis.  
 OS Synthetic.  
 XX  
 FT Key Location/Qualifiers  
 FT Region 109..128

FT  
 XX  
 PN CA2281913-A1. /label= B\_cell\_epitope  
 XX  
 PD 17-MAR-2001.  
 XX  
 PF 17-SEP-1999; 99CA-2281913.  
 XX  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 PA (KAYW/) KAY W.W.  
 PA (BURJ/) BURIAN J.  
 PA (KUZU/) KUZUK M.A.  
 XX  
 PI Kay WW, Burian J, Kuzuk MA;  
 XX  
 DR WPI: 2001-316844/34.  
 DR N-PSDB: AAF86247.  
 PT Method for protecting polkiothermic fish against salmonid rickettsial  
 PT septicemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the ospa protein of Piscirickettsia salmonis  
 XX  
 PS Disclosure; Fig 5; 35pp; English.  
 CC This invention relates to a method for the protection against infection  
 CC of a polkiothermic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic amount of a  
 CC salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC P. salmonis. The method is used for protecting animals,  
 CC Ospa in the form of a vaccine. The method is used for protecting against  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents optimised P. salmonis Ospa protein 17E2. The DNA  
 CC encoding Ospa 17E2 (AAF86247) has been optimised for expression in  
 CC Escherichia coli. An Ospa protein with an N-terminal fusion partner is  
 CC used in a vaccine to create an anti-Ospa antibody response.  
 CC  
 SQ Sequence 161 AA:  
 Query Match 62.3%; Score 832; DB 22; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-73;  
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 96 MRGCLGSSLLITSVFLVCAONFSROEVAATGAVVAGVAGQLFGKSGSRVSMAGVAV 155  
 DB 1 MRGCLGSSLLITSVFLVCAONFSROEVAATGAVVAGVAGQLFGKSGSRVSMAGVAV 60  
 QY 156 LGGLIGSKIGOSMDQDKIKLNSLEKVRAGVTRMRNPDGNSYSVEPVRYTORYNKOE 215  
 DB 156 LGGLIGSKIGOSMDQDKIKLNSLEKVRAGVTRMRNPDGNSYSVEPVRYTORYNKOE 120  
 QY 216 RROQYCFEPOOKAMIAQOKOETIGTACPOPGRMVOYISTEK 256  
 DB 121 RROQYCFEPOOKAMIAQOKOETIGTACPOPGRMVOYISTEK 161  
 RESULT 3  
 ID AAG78025 standard; Protein; 162 AA.  
 AC AAG78025;  
 XX 15-JAN-2002 (first entry)  
 DE Piscirickettsia salmonis polypeptide p10.6.  
 XX  
 KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;  
 KW septicemia; SRS; surface antigen; vaccine; antibacterial; fish;  
 KW AFCC VR-1361.  
 XX  
 OS Piscirickettsia salmonis.  
 OS

PN WO20016865-A2.  
PD 20-SEP-2001.  
XX  
XX  
XX  
XX 12-MAR-2001; 2001WO-CB01055.  
XX  
XX 11-MAR-2000; 2000GB--0005838.  
PR 01-JUL-2000; 2000GB--0016080.  
PR 01-JUL-2000; 2000GB--0016082.  
PR 29-JUL-2000; 2000GB--0018599.  
XX  
XX  
PA (AQUA-) AQUA HEALTH EURO LTD.  
XX  
XX Smaard N, Brouwers H, Jones S, Griffiths S, Valenzuela P,  
PI Burzio L;  
XX  
XX MPI: 2001-639050/73.  
DR N-PSDB; AAH79040.  
XX  
XX New nucleic acids encoding an amino acid sequence homologous to the  
PT surface antigen present on *Piscirickettsia salmonis* are useful to  
PT protect fish against *piscirickettsiosis* -  
XX  
XX Claim 6; Fig 5; 25pp: English.  
XX  
XX The invention relates to nucleic acid sequences and the encoded protein  
CC of a least part of the surface antigen present on *Piscirickettsia*  
CC salmonis for production of a vaccine with antibacterial activity to  
CC protect fish against *P. salmonis* which causes *piscirickettsiosis*, also  
CC known as salmonid rickettsial septicemia.  
XX  
XX Sequence 162 AA;

Query Match	Similarity	Score	DB	Length
Best Local	58.1%	Pred. No. 7.6e-72		162
Matches 158; Conservative	1; Mismatches 1; Indels 0; Gaps 0			
OY	97	RCLDGLSSLIITISVFLVGCAGNFNRQEVGAATGAVVGAAGOLFCKGSGRVSMAIGAVL	156	
DB	3	RCLDGLSSLIITISVFLVGCAGNFNRQEVGAATGAVVGAAGOLFCKGSGRVSMAIGAVL	62	
OY	157	GGLISGKIGQSMDDQDKIKLNGSLERKVGAGVTRMRNPDTGNSYSVEPVRTYQRYNKKOER	216	
DB	63	GGLISGKIGQSMDDQDKIKLNGSLERKVGAGVTRMRNPDTGNSYSVEPVRTYQRYNKKOER	122	
OY	217	RQCYCREFOQKAMIAQOKOETIGTACPPQPDGRMÖYISTEK	256	
DB	123	RQCYCREFOQKAMIAQOKOETIGTACPPQPDGRMÖYISTEK	162	
RESULT 4				
AAB81126				
ID	AAB81126	standard: Protein; 162 AA.		
XX	AAB81126;			
AC				
XX				
DT	11-JUL-2001	(first entry)		
XX				
DE				
XX	Ospa antigen amino acid sequence.			
KM	Polkilothemic fish; piscirickettsia salmonis; rickettsial pathogen;			
KW	Vaccine: Ospa; salmonid rickettsial septicaemia; rickettsial disease;			
XX	SRS.			
XX				
OS	piscirickettsia salmonis.			
FH	Key	Location/Qualifiers		
FT	Region	110..129		
FT		/label= B_cell_epitope		
XX				
PN	CA2281913-A1.			
XX				
PD	17-MAR-2001.			

XX PF 17-SEP-1999; 99CA-2281913.  
XX PR 17-SEP-1999; 99CA-2281913.  
XX PA (KAYW/) KAY M W.  
XX PA (BURJ/) BURIAN J.  
XX PA (KUZY/) KUZYK M A.  
XX PL Kay WW, Burian J, Kuzyk MA;  
XX WPI: 2001-316844/34.  
XX DR N-PDSB: AAF86246.  
PT Method for protecting poikilothermic fish against salmonid rickettsial  
PT septicemia and other rickettsial diseases comprises administering a  
XX vaccine containing the Ospa protein of Piscirickettsia salmonis .  
XX  
XX Example 2: Fig 2B; 35pp; English.

This invention relates to a method for the protection against infection of a poikilothermic fish by the bacterial pathogen, Piscirickettsia salmonis. The method comprises administering an immunogenic amount of a P. salmonis specific antigen termed ospa, or an immunogenic fragment of ospa in the form of a vaccine. The method is used for protecting animals, particularly poikilothermic fish, against the bacterial pathogen P. salmonis. The method is also useful for protecting against salmonid rickettsial septicemia (SRs) and other rickettsial diseases. The present sequence represents P. salmonis OspA protein. An OspA protein with an N-terminal fusion partner is used in a vaccine to create an anti-OspA antibody response.

Sequence 162 AA:

Query Match	Best Local Similarity	Score	DB	Length
Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0	98.8%;	815;	DB 22;	162;
			Pred. No. 7.6e-72;	
QY	97	RGCGGSSLIITISVFLVGCANFSRQEVGATGA	VYGVGAGQLFGKSGRYMAIGAVL	156
DB	3	RGCGGSSLIITISVFLVGCANFSRQEVGATGA	VYGVGAGQLFGKSGRYMAIGAVL	62
QY	157	GGILGKIGCSMDQDKIKLNOSLEKYKAGQVTRMR	PDGNSYSVPVRYORNRQER	216
DB	63	GGILGKIGCSMDQDKIKLNOSLEKYKAGQVTRMR	PDGNSYSVPVRYORNRQER	122
QY	217	ROOYCREFOOKAMTAGOKOEIVGTACPPDPGR	MOVISTEK	256
DB	123	ROOYCREFOOKAMTAGOKOEIVGTACPPDPGR	MOVISTEK	162
RESULT 5				
ID	AAE05746			
XX	AAE05746 standard; Protein; 190 AA.			
XX	AAE05746;			
XX	24-SEP-2001 (first entry)			
DE	Clostridium cellulovorans cellulose binding domain-180 (CBD-180).			
XX				
XX	Polysaccharide modification; polysaccharide binding domain; PBD; paper;			
KW	Yarn; fiber; textile; biological crosslinker; mechanical property;			
KW	Clostridium cellulovorans cellulose binding domain; wet strength;			
KW	durability; elasticity; CBDcos; cellulose binding protein A; CBP A;			
KW	CBD-180.			
XX				
OS	Clostridium cellulovorans.			
XX				
PN	WO200134091-A2.			
XX				
PD	17-MAY-2001.			
XX				

```

PF 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
XX 18-NOV-1999; 99US-0166389.
XX (CBPT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX Levy I, Nussinovitch A, Shoseyov O;
PI MPI: 2001-457121/49.
DR N-PSDB; AAD11043.
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX Example 1.2; Fig 1e-1g; 121pp; English.
PS The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the sizing step. The use of a biological crosslinker
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is Clostridium cellulovorans cellulose binding
CC domain (CBD)-180. CBD-180 contains 180 amino acids of C. cellulovorans
CC CBD (CBDC180) of cellulose binding protein A (CBP A).
XX SQ
Sequence 190 AA:
Query March 36.5%; Score 487; DB 22; Length 190;
Best Local Similarity 46.8%; Pred. No. 1.4e-39;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;
QY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQTFWCDHAGAL 60
DB 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQTFWCDHAGAL 60
QY 61 LGNSVYDNTSKYTANFVETASPTSTYDTPDPSHMRGLOGLSSLIISVFLVCAQMF 120
DB 61 LGNSVYDNTSKYTANFVETASPTSTYDTPDPSHMRGLOGLSSLIISVFLVCAQMF 120
QY 121 ROEVGAATGAVGVAGAGLFGKSGRVSMAIGAVLGLIGSKIQGSDQDQKIKLNSL 180
DB 93 -----FGFASGRATL-----KKQSFITIQ----- 111
QY 181 EKYKAGQVTR--WRNPDTGNSYSVEPRYQRYKNOERHQYCFREPOKAMLAGOKQET 238
DB 112 -----GRITKSDMSNTYTQNDYSFDSASSTPVVP-----KYTGITGAK--VL 153
QY 239 GTACPOPD 246
DB 154 GTR-PCPD 160

```

RESULT 6  
AAE05749  
ID AAE05749 standard; Protein; 328 AA.

AC AAE05749;  
XX 24-SEP-2001 (first entry)

```

XX Chimeric S peptide-cellulose binding domain-S protein.
DE Polysaccharide modification; polysaccharide binding domain; PBD; paper;
XX yarn; fiber; textile; biological crosslinker; mechanical property;
XX wet strength; durability; elasticity; cellulose binding domain; CBD;
XX chimeric protein; S peptide-cellulose binding domain-S protein;
XX Spep-CBD-Sprot; bovine.
OS Chimeric - Clostridium cellulovorans.
XX Chimeric - Bos sp.
XX Key Location/Qualifiers
XX Region 30..208
XX Region 226..326
XX Region /note="This region is derived from C. cellulovorans"
XX Misc-difference 327
XX /note="This region is derived from bovine"
XX /label="Unknown"
XX /note="Encoded by TAG"
XX Misc-difference 327..328
XX /note="These residues are absent in the sequence shown
XX as SEQ ID NO: 10 in the sequence listing"
XX WO200134091-A2.
XX 17-MAY-2001.
XX 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
XX 18-NOV-1999; 99US-0166389.
XX (CBPT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX Levy I, Nussinovitch A, Shoseyov O;
PI MPI: 2001-457121/49.
DR N-PSDB; AAD11046.
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX Example 4; Fig 4b-4g; 121pp; English.
PS The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles, using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the sizing step. The use of a biological crosslinker
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is S peptide-cellulose binding domain-S protein
CC (Spep-CBD-Sprot), a fusion protein derived from Clostridium cellulovorans
CC and bovine.
CC Note: This sequence is stated to be the same as that shown as
CC SEQ ID NO: 10 in the sequence listing of the specification. However
CC this sequence has 2 additional residues at its C-terminal.
XX SQ
Sequence 328 AA:
Query Match 36.5%; Score 487; DB 22; Length 328;

```



Best Local Similarity 46.8%; Pred. No. 2.9e-39;  
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

```

OY 1 MSVEFYNSKNSAQTNSTPTPIKITTNTSDSLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 60
DB 35 MSVEFYNSKNSAQTNSTPTPIKITTNTSDSLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 94
OY 61 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 120
DB 95 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 126
OY 121 ROEYGAATGAVGVAGVAGLFGKSGSRVSMAGAVLGLGSKIGSQMDQDKIKINQSL 180
DB 127 -----FGFASGRATL-----KKGQFTITQ----- 145
OY 181 EYKAKAGVTR--WRNPDTGNSYSVEPRYTORYNKORRQOYCREFOKAMIAQOKEIY 238
DB 146 ----GRTTSDMSNNTQTNDYSPDASSSTPVVNP-----KVTGYIGAK--VL 187
OY 239 GTACPOPD 246
DB 188 GTA-PCPD 194

```

RESULT 7  
AAE05747  
ID AAE05747 standard; protein; 341 AA.

AAE05747;

24-SEP-2001 (first entry)

Clostridium cellulovorans CBD cross linker protein (CCP).

Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
yarn; fiber; textile; biological crosslinker; mechanical property;  
wet strength; durability; elasticity; cellulose binding domain; CBD;  
CBD cross linker protein; CCP.

Clostridium cellulovorans.

Key Location/Qualifiers

Misc-difference 340..341

/note- "Encoded by CCATAGAT"

Misc-difference 341

/note- "This residue is absent in the sequence shown  
as SEQ ID NO: 6 in the sequence listing"

W0200134091-A2.

17-MAY-2001.

02-NOV-2000; 2000MO-IL00708.

08-NOV-1999; 99US-0164140.

18-NOV-1999; 99US-0166389.

(CBOT-) CBD TECHNOLOGIES LTD.

(YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

Levy I, Nussinovitch A, Shoseyov O;

WPI: 2001-457121/49.

N-PDB; AAD11044.

Preparation of a polysaccharide containing material having at least one  
desired structural, chemical, physical, electrical and/or mechanical  
property

Example 1.2; Fig 2b-2e; 121pp; English.

The present invention relates to methods and compositions for cross-  
linking and/or modifying the properties of polysaccharide materials.

The method involves treating the polysaccharide structure with a  
CC polysaccharide binding domain (PBD) fusion protein. The method is  
used to alter the structural, chemical, physical, electrical and  
mechanical properties of polysaccharide materials such as paper,  
CC yarns, fibers and textiles, using biological crosslinking agents.  
The polysaccharide-containing materials have improved mechanical  
CC properties such as wet strengths, durability and elasticity. The PBD  
reagent is applied in the forming stage in fluting paper manufacture  
CC which eliminates the sizing step. The use of a biological crosslinker  
improves the recyclability of paper products. The PBD reagent maintains  
the fine fibers in a slurry therefore resulting in better recovery of  
CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
CC which enhances the ability of the alkaline glue used in binding to  
penetrate paper.

The present sequence is Clostridium cellulovorans cellulose binding  
domain (CBD) cross linker protein (CCP).

Note: This sequence is stated to be the same as that shown as  
CC SEQ ID NO: 6 in the sequence listing of the specification. However  
this sequence has an additional residue at its C-terminal.

Sequence 341 AA:

Query Match 36.5%; Score 487; DB 22; Length 341;

Best Local Similarity 46.8%; Pred. No. 3.1e-39;  
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

```

OY 1 MSVEFYNSKNSAQTNSTPTPIKITTNTSDSLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 60
DB 1 MSVEFYNSKNSAQTNSTPTPIKITTNTSDSLNLDVAVRRYRYSYSDGQGTFCMDHAGAL 60
OY 61 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 120
DB 61 LGNSYVDNTSKVTANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVCAQNF 120
OY 121 ROEYGAATGAVGVAGVAGLFGKSGSRVSMAGAVLGLGSKIGSQMDQDKIKINQSL 180
DB 121 ROEYGAATGAVGVAGVAGLFGKSGSRVSMAGAVLGLGSKIGSQMDQDKIKINQSL 180
OY 93 -----FGFASGRATL-----KKGQFTITQ----- 111
DB 93 -----FGFASGRATL-----KKGQFTITQ----- 111
OY 181 EYKAKAGVTR--WRNPDTGNSYSVEPRYTORYNKORRQOYCREFOKAMIAQOKEIY 238
DB 112 ----GRTTSDMSNNTQTNDYSPDASSSTPVVNP-----KVTGYIGAK--VL 153
OY 239 GTACPOPD 246
DB 154 GTA-PCPD 160

```

RESULT 8  
AAR63634  
ID AAR63634 standard; protein; 162 AA.

AAR63634;

05-JUN-1995 (first entry)

Cellulose binding domain.

Cellulose binding domain; CBD.

Clostridium cellulovorans.

W09424158-A.

27-OCT-1994.

14-APR-1994; 94MO-US04132.

14-APR-1993; 93US-0048164.

(REGC) UNIT CALIFORNIA.

(YISS) YISSUM RES & DEV CO.

Dol RH, Goldstein MA, Shoseyov O, Shpiegel I;

XX WPI: 1994-341767/42.  
 DR N-PSDB; AA072917.  
 XX Isolated cellulose binding domain and fusion proteins - with  
 PT applications, including drug delivery, affinity separations, and  
 PT diagnostic techniques  
 XX Claim 1; Fig 1; 125pp; English.  
 PS  
 XX AA072917/R63634 is a novel isolated cellulose binding domain. It pref.  
 CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,  
 CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and  
 CC the encoding DNA are claimed, as is a CBD fusion protein (FP)  
 CC comprising the CBD and a second protein. The second protein is pref.  
 CC Protein A, heparinase, a hormone or an enzyme capable of degrading  
 CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive  
 CC HSP-related protein or an antigenic portion of this. The CBD and FP  
 CC may be used in drug delivery, affinity seps. and diagnostic  
 CC techniques. CBD nucleic acid may be obtd. from a variety of cell  
 CC sources that produce CBDs that bind with high affinity and in a  
 CC reversible manner or that produce CBD encoding mRNA. The preferred  
 CC source of CBD encoding nucleic acid is C. cellulovorans.  
 CC  
 SQ Sequence 162 AA:  
 Query Match 36.2%; Score 483; DB 15; Length 162;  
 Best Local Similarity 51.7%; Pred. No. 2.7e-39; Mismatches 9; Indels 14; Gaps 4;  
 Matches 106; Conservative 9;  
 QY 1 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 60  
 Db 6 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 65  
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPSHMRGCLQSSLIISVFLVCAQNF 120  
 Db 66 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPSHMRGCLQSSLIISVFLVCAQNF 97  
 QY 121 ROEYGAATGAVGVAGVAGQLFKSGSGRVSMAGAVLGLGSKIGSQMDQDKIKLNSL 180  
 Db 98 -----FGFASGRATL-----KKGQFTITQ----- 116  
 QY 181 EKVKAGQVTR--WRNPDTGNSYSVE 203  
 Db 117 -----GRITKSDMSNNTYTNDYSFD 136  
 RESULT 9  
 ID AAE05745 standard; Protein: 163 AA.  
 XX AAE05745;  
 AC 24-SEP-2001 (first entry)  
 XX  
 DE Clostridium cellulovorans cellulose binding domain (CBDClos).  
 XX  
 XX Clostridium cellulovorans cellulose binding domain; PBD; paper;  
 KW yarn; fiber; textile; biological crosslinker; mechanical property;  
 KW Clostridium cellulovorans cellulose binding domain; wet strength;  
 KW durability; elasticity; CBDClos; cellulose binding protein A; CBP A.  
 XX  
 XX Clostridium cellulovorans.  
 OS  
 XX WO200134091-A2.  
 PN  
 XX 17-MAY-2001.  
 PD  
 XX 02-NOV-2000; 2000WO-IL00708.  
 PF  
 XX 08-NOV-1999; 99US-0164140.  
 PR 18-NOV-1999; 99US-0166389.  
 XX

PA (CBDT-) CBD TECHNOLOGIES LTD.  
 PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
 XX  
 XX Levy I, Nussinovitch A, Shoseyov O;  
 PI  
 XX WPI: 2001-457121/49.  
 DR N-PSDB; AA011042.  
 XX Preparation of a polysaccharide containing material having at least one  
 PT desired structural, chemical, physical, electrical and/or mechanical  
 PT property  
 XX  
 XX Example 1.1: Page 111-112; 121pp; English.  
 PS  
 XX The present invention relates to methods and compositions for cross-  
 XX linking and/or modifying the properties of polysaccharide materials.  
 CC The method involves treating the polysaccharide structure with a  
 CC polysaccharide binding domain (PBD) fusion protein. The method is  
 CC used to alter the structural, chemical, physical, electrical and  
 CC mechanical properties of polysaccharide materials such as paper,  
 CC yarns, fibers and textiles, using biological crosslinking agents.  
 CC The polysaccharide containing materials have improved mechanical  
 CC properties such as wet strengths, durability and elasticity. The PBD  
 CC reagent is applied in the forming stage in futing paper manufacture  
 CC which eliminates the sizing step. The use of a biological crosslinker  
 CC improves the recyclability of paper products. The PBD reagent maintains  
 CC the fine fibers in a slurry therefore resulting in better recovery of  
 CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
 CC which enhances the ability of the alkaline glue used in binding to  
 CC penetrate paper.  
 CC The present sequence is Clostridium cellulovorans cellulose binding  
 CC domain (CBDClos) of cellulose binding protein A (CBP A).  
 CC  
 SQ Sequence 163 AA:  
 Query Match 36.2%; Score 483; DB 22; Length 163;  
 Best Local Similarity 51.7%; Pred. No. 2.7e-39; Mismatches 14; Indels 76; Gaps 4;  
 Matches 106; Conservative 9;  
 QY 1 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 60  
 Db 7 MSVEFNSKNSAQTNSITPIIKITNTSDSLNDLVKRYRYTSDGTGQGFPCDHAGAL 66  
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPSHMRGCLQSSLIISVFLVCAQNF 120  
 Db 67 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPSHMRGCLQSSLIISVFLVCAQNF 98  
 QY 121 ROEYGAATGAVGVAGVAGQLFKSGSGRVSMAGAVLGLGSKIGSQMDQDKIKLNSL 180  
 Db 99 -----FGFASGRATL-----KKGQFTITQ----- 117  
 QY 181 EKVKAGQVTR--WRNPDTGNSYSVE 203  
 Db 118 -----GRITKSDMSNNTYTNDYSFD 137  
 RESULT 10  
 ID AAE05748 standard; Protein: 428 AA.  
 XX AAE05748;  
 AC 24-SEP-2001 (first entry)  
 XX  
 DE Clostridium cellulovorans protein A-cellulose binding domain.  
 XX  
 XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
 KW yarn; fiber; textile; biological crosslinker; mechanical property;  
 KW wet strength; durability; elasticity; protein A-cellulose binding domain;  
 KW ProDA-CBD.  
 XX  
 XX Clostridium cellulovorans.  
 OS  
 XX

Oy	121	KQEVGAATGAVGVCAQGLFGKSGSRVSAIGCAVLGCLISGICQSDMDQDKIKLNLSL	180
Dd	362	-----FGFASGRATL-----KKCFITID-----	380
Oy	181	EKVAKGOVTR--WRNPDTGNSYSVE	203
Dd	381	----GRTTKSDMSNVTQTNDYSPD	400
		RESULT 11	
ID	AAY39952		
AC	AAY39952 standard; Protein; 382 AA.		
XX	AAY39952;		
DT	13-DEC-1999 (first entry)		
DE	Gaussia luciferase fusion protein sequence.		
XX	Luciferase; green fluorescent protein; GFP; screening assay; diagnosis;		
KW	bioluminescence-generating system; toy; cosmetic; fairy dust; beverage;		
RW	body paint; squirt gun; balloon; slimy play material; soap; toothpaste;		
KM	fusion protein.		
XX			
OS	Gaussia sp.		
XX			
PN	WO9949019-A2.		
PD	30-SEP-1999.		
XX			
PF	26-MAR-1999; 99MO-U0506698.		
XX			
PR	27-MAR-1998; 98US-0079624.		
PR	15-JUN-1998; 98US-0089367.		
PR	01-OCT-1998; 98US-0102939.		
XX			
PA	(PROL-) PROLUME LTD.		
PA	(BRYA/) BRYAN B. J.		
XX			
PI	Bryan BJ, Szent-Gyorgyi C;		
DR	WP1: 1999-580443/49.		
XX	N-PSDB: AAZ27550.		
PT	New isolated Renilla mulleri, Gaussia and Pleuromamma luciferase and		
PT	Renilla and Ptilosarcus green fluorescent protein nucleic acids -		
PS	Disclosure: Page 222-223; 233pp; English.		
CC	This sequence represents a luciferase of the invention. The invention		
CC	relates to Renilla mulleri, Gaussia and Pleuromamma luciferase and		
CC	Renilla and Ptilosarcus green fluorescent protein (GFP) nucleic acids and		
CC	proteins. The luciferases and GPPs can be used in		
CC	bioluminescence-generating systems, assays, screening methods, diagnostic		
CC	method and articles of manufacture. They can be expressed using		
CC	e.g. bacterial, yeast, fungal, plant, insect or animal cells. The Renilla		
CC	mulleri, Gaussia and Pleuromamma luciferase or Renilla or Ptilosarcus		
CC	GPPs can be used in e.g. toys, cosmetics, fountains, personal care items,		
CC	fairy dust, beverages, soft drinks, foods, textile products, bubbles,		
CC	balloons, personal items, denturifrices, soaps, body paints, bubble bath,		
CC	ink or paper products. In particular, they can be used in e.g. squirt		
CC	guns, pellet guns, finger paints, foot bags, greeting cards, slimy play		
CC	material, clothing, bubble making toys, bath powders, cosmetics, body		
CC	lotions, gels, body powders, body creams, toothpastes, mouthwashes,		
CC	soaps, body paints, bubble bath, inks, wrapping paper, gelatins, icings,		
CC	frostings, greeting cards, beer, wine, champagne, soft drinks, ice cubes		
CC	ice, dry ice or fountains. The nucleic acids can also be used to produce		
CC	transgenic fish and plants.		
XX			
SO	Sequence 382 AA;		
Query Match	36.0%	Score 480.5; DB 20; Length 382;	
Best Local Similarity	62.6%;	Pred. No. 1,6e-38;	

Matches 107; Conservative 10; Mismatches 33; Indels 21; Gaps 4;

QY 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKKVRYRTSDGTGOTFWCHAGAL 60  
 DB 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKKVRYRTSDGTGOTFWCHAGAL 60  
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYFLDPSHMRGCL---QGSSLIITISVFLVCAQ 117  
 DB 61 LGNSYVDNTSKVTANFVKETASPTSTYDYFLDPSHMRGCL---QGSSLIITISVFLVCAQ 120  
 QY 118 NESRO-----EVGAATGAVGVAGOLFGKSGRVSNAIGCA-VLGGILGS 162  
 DB 121 NTQVINDSFPDASSSTPYVNPKYTG-----IGGAKVLTGARGS 159

## RESULT 12

AAM90077 standard; Protein: 162 AA.

AC AAM90077;

DT 09-MAR-1999 (first entry)

DE C. cellulovorans Cbpa CBD protein.

XX Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;  
 KW immunoassay; heat-shock protein; cross reactive protein; detection;  
 KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;  
 KM cellulose; chitin; cellulolytic; amorphogenic.

XX Clostridium cellulovorans.

XX US5856201-A.

PD 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

PR 14-APR-1993; 93US-0048164.

PR 14-APR-1994; 94WO-US04132.

XX (REGC ) UNIV CALIFORNIA.

PA (YISS ) YISSUM RES & DEV CO.

PI Doi RH, Goldstein MA, Shoseyov O, Shplegl I;

DR WPI; 1999-105130/09.

XX N-PSDB; AAV74072.

XX Detection of a specific analyte by reaction with binding agent fused

PT to cellulose binding domain - and subsequent treatment with

PT cellulose and reaction of insoluble product with a label specific

PT for the analyte

XX Claim 3; Fig 1A-B; 63pp; English.

XX This sequence represents a cellulose binding domain (CBD) derived from

CC the Cbpa protein of Clostridium cellulovorans. The sequence is used the

CC construction of a fusion protein which can be used in diagnostic

CC immunoassays, e.g. to detect heat-shock proteins (HSP) and their

CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies

CC (which indicate insulin-dependent diabetes mellitus, or susceptibility

CC to it. The CBD, has very high affinity for cellulose (including

CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but

CC has almost no cellulolytic or amorphogenic activities. The CBD binds over

CC a wide pH range and is not released from cellulose by washing with water.

XX Sequence 162 AA;

SQ

Query Match 35.7%; Score 477; DB 20; Length 162;

Best Local Similarity 97.8%; Pred. No. 1e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 90; Conservative

QY 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKKVRYRTSDGTGOTFWCHAGAL 60  
 DB 6 MSVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKKVRYRTSDGTGOTFWCHAGAL 65

QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYFLD 92

DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYFLD 97

## RESULT 13

AAM90081 standard; Protein: 154 AA.

AC AAM90081;

DT 09-MAR-1999 (first entry)

DE C. cellulovorans CBD-Kpni fusion protein.

XX Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;  
 KW immunoassay; heat-shock protein; cross reactive protein; detection;  
 KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;  
 KM cellulose; chitin; cellulolytic; amorphogenic.

XX Clostridium cellulovorans.

XX Synthetic.

XX US5856201-A.

PD 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

PR 14-APR-1993; 93US-0048164.

PR 14-APR-1994; 94WO-US04132.

XX (REGC ) UNIV CALIFORNIA.

PA (YISS ) YISSUM RES & DEV CO.

PI Doi RH, Goldstein MA, Shoseyov O, Shplegl I;

DR WPI; 1999-105130/09.

XX Detection of a specific analyte by reaction with binding agent fused

PT to cellulose binding domain - and subsequent treatment with

PT cellulose and reaction of insoluble product with a label specific

PT for the analyte

XX Claim 5; Column 55-56; 63pp; English.

XX This sequence represents a fusion protein (CBD-Kpni) composed of the Cbpa

CC protein cellulose binding domain (CBD) from Clostridium cellulovorans

CC This protein is used in diagnostic immunoassays, e.g. to detect

CC heat-shock proteins (HSP) and their cross-reactive proteins, antigenic

CC fragments or HSP-specific antibodies (which indicate insulin-dependent

CC diabetes mellitus, or susceptibility to it. The CBD has very high

CC affinity for cellulose (including crystalline forms) and chitin

CC (dissociation constant 0.8-1.4 mu M), but has almost no cellulolytic or

CC amorphogenic activities. The CBD binds over a wide pH range and is not

CC released from cellulose by washing with water.

XX Sequence 154 AA;

SQ

Query Match 35.4%; Score 472; DB 20; Length 154;

Best Local Similarity 97.8%; Pred. No. 3e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 89; Conservative

QY 2 SVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKKVRYRTSDGTGOTFWCHAGAL 61

DB 1 SVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKKVRYRTSDGTGOTFWCHAGAL 60

OY 62 GNSYVDNTSKVTANFYKETASPTSTYDTYLD 92  
 DB 61 GNSYVDNTSKVTANFYKETASPTSTYDTYVE 91

## RESULT 14

AAW90080  
 ID AAW90080 standard; Protein: 156 AA.

XX AAW90080;

AC AAW90080;

XX 09-MAR-1999 (first entry)

DE C. cellulovorans CBD-HSP fusion protein.

KM Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;

KW Immunassay; heat-shock protein; cross reactive protein; detection;

KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;

XX cellulose; chitin; cellulolytic; amorphogenic.

OS Clostridium cellulovorans.

XX Synthetic.

PN US5856201-A.

XX 05-JAN-1999.

PD 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

XX 14-APR-1993; 93US-0048164.

PR 14-APR-1994; 94WO-US04132.

XX (REGC ) UNITV CALIFORNIA.

PA (YISS ) YISSUM RES & DEV CO.

XX Dol RH, Goldstein MA, Shoseyov O, Shplegl I;

PI WPI: 1999-105130/09.

XX WPI: 1999-105130/09.

DR Detection of a specific analyte by reaction with binding agent fused

PT to cellulose binding domain - and subsequent treatment with

PT cellulose and reaction of insoluble product with a label specific

PT for the analyte

XX Claim 4: Column 51-52; 63pp; English.

PS This sequence represents a fusion protein composed of the Cbpa protein

CC cellulose binding domain (CBD) from Clostridium cellulovorans and a heat

CC shock protein (HSP) fragment. This protein is used in diagnostic

CC immunassays, e.g. to detect heat-shock proteins (HSP) and their

CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies

CC to it. The CBD has very high affinity for cellulose (including

CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but

CC has almost no cellulolytic or amorphogenic activities. The CBD binds over

CC a wide pH range and is not released from cellulose by washing with water.

XX Sequence 156 AA;

Query Match 35.4%; Score 472; DB 20; Length 156;

Best Local Similarity 97.8%; Pred. No. 3e-38; Indels 0; Gaps 0;

Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SVEFYNSNKSQAQNTSTPIIKITNTSDSLNDVVRVRYTSDGOGTFMCDHAGALL 61

DB 1 SVEFYNSNKSQAQNTSTPIIKITNTSDSLNDVVRVRYTSDGOGTFMCDHAGALL 60

OY 62 GNSYVDNTSKVTANFYKETASPTSTYDTYLD 92

DB 61 GNSYVDNTSKVTANFYKETASPTSTYDTYVE 91

RESULT 15  
 AAR95080  
 ID AAR95080 standard; peptide: 167 AA.

XX AAR95080;

AC AAR95080;

XX 30-OCT-1996 (first entry)

DE Cellulose binding domain.

KM Cellulose binding domain; CBD; hapten; moiety; biotin; avidin;

KW streptavidin; affinity chromatography; cell separation;

KM cell immobilisation; protein immobilisation; enzyme immobilisation;

KW multienzyme reactors; signal immunassays; drug delivery; pesticide;

XX cellulose; chitin.

OS Clostridium thermocellum.

XX Key Location/Qualifiers

FT Misc-difference 139 /note= "Unidentified amino acid."

XX WO9613524-A1.

XX 09-MAY-1996.

PD 26-OCT-1995; 95WO-US13813.

XX 27-OCT-1994; 94IL-0111415.

XX (TECR ) TECHNION RES & DEV FOUND LTD.

PA (UYRA-) UNITV RAMOT APPLIED RES & IND DEV LTD.

PA (YEDA ) YEDA RES & DEV CO LTD.

XX Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;

PI WPI: 1996-239453/24.

XX WPI: 1996-239453/24.

DR Modified cellulose-binding domain moles - having attached hapten,

PT partic. biotin, to provide an affinity system for, e.g. seps.,

PT assays, reactors, delivery etc.

XX Claim 12: Page 34-35; 53pp; English.

PS A modified cellulose binding domain (CBD) or fraction of it, linked

CC to a hapten moiety via one or more Cys or Lys residues and complexes

CC comprising the biotinylated CBD and a biotin-binding molecule

CC selected from modified or unmodified avidin or streptavidin or an

CC anti-biotin antibody, can be used in affinity chromatography, cell

CC separation, cell, protein and enzyme immobilisation, selective

CC removal of biological materials, multienzyme reactors, signal

CC immunassays and drug (e.g. pesticide) delivery. The CBD molecule

CC can be modified with the hapten without affecting its high affinity

XX for cellulose, or chitin.

XX Sequence 167 AA;

Query Match 18.4%; Score 245; DB 17; Length 167;

Best Local Similarity 51.5%; Pred. No. 6e-16;

Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

OY 1 MSVEFYNSNKSQAQNTSTPIIKITNTSDSLNDVVRVRYTSDGOGTFMCDHAGALL 60

DB 9 LKVEFYNSNKSQAQNTSTPIIKITNTSDSLNDVVRVRYTSDGOGTFMCDHAGALL 67

OY 61 LCN--SYVDNTSKVTANFYKETASPTSTYDTYLDPSHMRCLQ 101

DB 68 ICSNGSYNGTITSNVKGTFVKMSSS-TNNADTYLEISTGTGLE 109

Search completed: October 27, 2002, 10:57:58  
 Job time : 36.487 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:01 ; Search time 28.7392 Seconds  
(without alignments)  
1540.986 Million cell updates/sec

Title: US-09-677-374-6  
 Perfect score: 1335  
 Sequence: 1 MSVEFYNSNKSQNTSITPI.....ITYGTACPPQDGRWQVISTEK 256

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 5622222

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum match 0%
	Maximum Match 100%
	Listing first 45 summaries

```

1:  sp.archaea: *
2:  sp.bacteria: *
3:  sp.fungi: *
4:  sp.insecta: *
5:  sp.invertebrate: *
6:  sp.mammal: *
7:  sp.mhc: *
8:  sp.ornafelle: *
9:  sp.phage: *
10: sp.plant: *
11: sp.rodent: *
12: sp.virus: *
13: sp.vetbrate: *
14: sp.unclassified: *
15: sp.rvirus: *
16: sp.bacteriap: *
17: sp.archaeop: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	length	DB	ID	Description
1	815	61.0	162	2	09f9x8		09f9x8 pisciricket
2	303.5	22.7	148	2	054381		054381 rickettsia
3	283	21.2	159	2	09f9f2		09f9f2 rickettsia
4	258.5	19.4	137	2	052252		052252 rickettsia
5	252.5	18.9	144	2	09k2n6		09k2n6 male-killin
6	251.5	18.8	137	2	031065		031065 rickettsia
7	251.5	18.8	144	2	09k4w8		09k4w8 male-killin
8	249	18.7	1546	2	045596		045596 clostridium
9	244	18.3	154	2	053154		053154 rickettsia
10	239	17.9	151	2	09f909		09f909 rickettsia
11	236.5	17.7	131	2	09f001		09f001 rickettsia
12	236.5	17.7	131	2	052637		052637 rickettsia
13	235.5	17.6	131	2	09f522		09f522 rickettsia
14	233.5	17.5	1162	2	082830		082830 clostridium
15	223.5	16.7	1483	16	09777y4		09777y4 clostridium
16	216.5	16.2	105	2	031208		031208 rickettsia

17	195.5	14.6	191.5	2	Q9RPJ0	Q9RPJ0	aceticivibrium
18	193	14.5	88.7	2	Q9L3J8	Q9L3J8	clostridium
19	169.5	12.7	3.21	2	Q46J92	Q46J92	clostridium
20	152.5	11.4	99.7	2	Q9Z4I1	Q9Z4I1	bacillus sp
21	151.5	11.3	177.9	2	Q52374	Q52374	caldicellul
22	150.5	11.3	177.0	2	Q9X3P5	Q9X3P5	caldicellul
23	149.5	11.2	26.1	2	Q9AGC7	Q9AGC7	caldicellul
24	149.5	11.2	14.26	2	Q9X3P6	Q9X3P6	caldicellul
25	148.5	11.1	17.51	2	Q9AGC4	Q9AGC4	caldicellul
26	144.5	10.8	98.6	2	Q9A0H0	Q9A0H0	caldicellul
27	142.5	10.7	1.000	2	Q24E20	Q24E20	thermophil
28	139	10.6	171.1	2	P96311	P96311	anaerocellu
29	132	10.4	7.7	2	Q9ACJ7	Q9ACJ7	ricketsia
30	138.5	10.4	9.21	2	Q9L8L8	Q9L8L8	caldicellul
31	137	10.3	1.59	16	Q9B5G4	Q9B5G4	rhizobium
32	135.5	10.1	9.30	2	Q9RXP5	Q9RXP5	caldicellul
33	128	9.6	1.70	2	Q9RFX6	Q9RFX6	caldicellul
34	127.5	9.6	18.2	16	Q9HX13	Q9HX13	pseudomon
35	122	9.1	2.57	16	Q9A8M8	Q9A8M8	caulobacte
36	121	9.1	4.99	2	Q93FJ6	Q93FJ6	bacillus su
37	121	9.1	5.08	2	Q93LD0	Q93LD0	bacillus su
38	120.5	9.0	1.36	16	Q9ZRH9	Q9ZRH9	rhizobium
39	114.5	8.6	5.01	2	Q830J2	Q830J2	caldicellul
40	114	8.5	4.99	2	Q52J31	Q52J31	caldicellul
41	113.5	8.5	4.99	2	Q45532	Q45532	caldicellul
42	111.5	8.4	15.4	16	Q914S1	Q914S1	pseudomon
43	111.5	8.4	23.16	2	Q9FNU9	Q9FNU9	bacteroides
44	111	8.3	1.55	2	Q9FEB1	Q9FEB1	edwardsiell
45	110	8.2	1.55	2	Q9RB08	Q9RB08	pectobacter

## ALIGNMENTS

## RESULT 1

ID Q9F9K8 PRELIMINARY; PRT; 162 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE 17 KDA ANTIGEN.  
GN OSPA.  
OS *Piscirickettsia salmonis*.  
OC Bacteria; Proteobacteria; gamma subdivision; *Piscirickettsia* group;  
OC *Piscirickettsia*.  
OX NCBI\_TaxID=1238;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LF-89;  
RA Kuznyk M.A., Burian J., Thornton J.C., Kay W.W.;  
RT "Identification of a genus-common *Rickettsia* surface antigen in the  
RL salmonid pathogen *Piscirickettsia salmonis*."; Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
EMBL: AF184152; AAC17000.1; -  
SQ SEQUENCE 162 AA: 17661 MW: DE9936FD94A527E CRC64;

Query Match 61.0%; Score 815; DB 2; Length 162;

QY	97	RGLDGLSSLIILISFLVACCAONFSQOEKATGAVGVAGOLFPGKSGSRVMAIGAVL	156
Db	3	RGLDGLSSLIILISFLVACCAONFSQOEKATGAVGVAGOLFPGKSGSRVMAIGAVL	62
QY	157	GGIISKIGQSMDDODKIKLWQSLKRYKAGVTRMRNPDDTGNSTSVPEPRVYQYNNQER	216
Db	63	GGIISKIGQSMDDODKIKLWQSLKRYKAGVTRMRNPDDTGNSTSVPEPRVYQYNNQER	122
QY	217	POOYCRFEQKAMTAGKQOEIYGTACPPDPGRMVISTEK	256
Db	123	POOYCRFEQKAMTAGKQOEIYGTACPPDPGRMVISTEK	162

```

RESULT 2
054381 ID PRELIMINARY; PRT: 148 AA.
AC 054381;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE 17 KDA COMMON-ANTIGEN (FRAGMENT).
OS Rickettsia sp.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=789;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98087556; PubMed=9425244;
RA Davis M.J., Ying Z., Brunner B.R., Pantofia A., Ferwerda F.H.;
RT "Rickettsial relative associated with papaya bunchy top disease.";
RL Curr. Microbiol. 36:80-84(1998).
DR EMBL; U76907; AAC02809.1; -.
FT NON_TER 1
FT SEQUENCE 148 AA; 15050 MW; A7AFEEFDE0AEED4C CRC64;

Query Match 22.7%; Score 303.5; DB 2; Length 148;
Best local similarity 40.1%; Pred. No. 2.2e-17;
Matches 55; Conservative 30; Mismatches 43; Indels 9; Gaps 2;

QY 120 SROEVGATGAVVGVAGOLFEGKSGRVSMATGAVLGLGSKRTGSGMDQDK----IK 175.
Db 17 NKGGSGTIGLIGLGVSGFGGGRGLAVAGALLGLNQLAGMDQDKRLALT 76
QY 176 LMSLEKRVAGVTRMNPDTGNSVPEVRYRQYRNKQEROQYCRFQOKAMTAGQK 235
Db 77 SQRLKAPSGSSVQMPNPNNGNYGVTPSKAV-----KNNNGYQCREYTYTVVVGKQ 131
QY 236 EYGTACPOPDGRWOV 252
Db 132 KAYGTACROPDGMQOV 148

RESULT 3
09F9F2 ID PRELIMINARY; PRT: 159 AA.
AC 09F9F2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 17 KDA GENUS-COMMON ANTIGEN.
OS Rickettsia felis (Rickettsia azadi).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=42862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21217364; PubMed=11321078;
RA Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,
RA Zavala-Velazquez J.E., Foil L.D., Stothard D.R., Azad A.F.,
RA Walker D.H.;
RT "Rickettsia felis: molecular characterization of a new member of the
RT spotted fever group.";
RL Int. J. Syst. Evol. Microbiol. 51:339-347(2001).
DR EMBL; AF195118; AAC28452.1; -.
SQ SEQUENCE 159 AA; 16457 MW; 34C5B020AFA70A1F CRC64;

Query Match 21.2%; Score 283; DB 2; Length 159;
Best local similarity 37.9%; Pred. No. 1.1e-15;
Matches 61; Conservative 30; Mismatches 34; Indels 16; Gaps 6;

QY 103 SLLIISV--FLVGC--AQNFRGVGATGAVVGVAGOLFEGKSGRVSMATG- GAVL 156
Db 5 SKMIITLASMILQACNGPGMKNKGCTGTLGAGCALLGSGFGKKGQL-VGVVGALL 63

```

```

QY 157 GGLIGSKIGSMQDQK----IKLMSLEKRVAGVTRMNPDTGNSVPEVRYRQYN 212
Db 64 GAVLGGIGAGMDEODRRLAELTSGRALETPDSGTISVEKRNDRNHNHGYVPNKTY----- 119
QY 213 KQHRROOYCRFQOKAMTAGOKOEIYGTACPOPDGRWOV 253
Db 120 -RNSTGQYCREYTYTVVVGKQKAYGNACROPDGMQOVN 159

RESULT 4
052252 ID PRELIMINARY; PRT: 137 AA.
AC 052252;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 17 KDA ANTIGEN (FRAGMENT).
OS Rickettsia cooley.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=69410;
RN [1]
RP SEQUENCE FROM N.A.
RX Billings A.N., Telow G.J., Walker D.H.;
RT "Molecular characterization of a novel spotted fever group rickettsial
RT species from ixodes scapularis in Texas.";
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF031534; AAB95267.1; -.
FT NON_TER 1
FT SEQUENCE 137 AA; 14215 MW; A27597A9FDB85FC3 CRC64;

Query Match 19.4%; Score 258.5; DB 2; Length 137;
Best local similarity 39.3%; Pred. No. 9.4e-14;
Matches 53; Conservative 25; Mismatches 46; Indels 11; Gaps 4;

QY 117 QNRSROEVGATGAVVGVAGOLFEGKSGRVSMATG- GAVLGLGSKRTGSGMDQDK-- 173
Db 7 RGMKQGTGTLGAGCALLGSGFGKQQL-VGVVGALLGAVLGGQIGAGMDEQDKRL 65
QY 174 --IKLMSLEKRVAGVTRMNPDTGNSVPEVRYRQYRNKQEROQYCRFQOKAMIA 231
Db 66 AELTSGRALPAASGSGSTEWNRPNNGNYGVTPNKTY-----RNSTGQYCREYTYTVIG 120
QY 232 GQKOEIYGTACPOPD 246
Db 121 GKQKAYGNACROPD 135

RESULT 5
09K2N6 ID PRELIMINARY; PRT: 144 AA.
AC 09K2N6;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE 17KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).
OS male-killing Rickettsia from Adalia bipunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=38028;
RN [1]
RP SEQUENCE FROM N.A.
RX Schlenker H.J.G.V.D., Habig M., Stoggett J.J., Webberley M.K.,
RA Bertrand D., Hurst G.D.D., Majerus M.E.N.;
RT "On the evolution of male-killing: Monophyletic origin and horizontal
RT transfer of male-killing Rickettsia (a-Proteobacteria) from two con-
RT generic ladybirds, Adalia bipunctata L. and A. decempunctata L.
RT (Coleoptera: Coccinellidae).";
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ269518; CAB96383.1; -.
DR EMBL; AJ269517; CAB96382.1; -.
FT NON_TER 1
FT SEQUENCE 144 144

```



SEQUENCE	144 AA:	14785 MW:	CB254739CCAA56AE7	CRC64:
SO	SEQUENCE	144 AA:	14785 MW:	CB254739CCAA56AE7
Query Match	18.9%:	Score 252.5:	DB 2:	Length 144:
Best Local Similarity	39.4%:	Pred. No. 3.1e-13:		
Matches	52:	Conservative	24:	Mismatches 45; Indels 11; Gaps 4;
OY	120	SNQEVGATGAVGVGAQGLPFKSGGRVSMATG-CAVJLGLGSKTIGSGMDQDK----	174	
DB	17	KNQGTGTLLGGAGGALLSGQFGKGQL-VGVGVALLGAVGGGIGAGMDQDRRLAEL	75	
OY	175	KLNQSLLEKAKAQVRRMRPDTGNSYSVEPVRTYORYNKOEPRQOYCREFOQKAMIAQOK	234	
DB	76	TSQRALLEAPSSSNVEMRRPDMGNNGYTPNKTY-----RNSTGYCRREYTOTVIYIGKQ	130	
OY	235	QEIYGTACPPD	246	
DB	131	OKAYGNACRQPD	142	
RESULT 6				
OJ1065	PRELIMINARY:	PRT:	137 AA:	
AC	OJ1065:	OJ9M02:		
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-OCT-2001	(TREMBLrel. 18, Last annotation update)		
DE	17	KDA ANTIGEN (17 KDA PROTEIN) (FRAGMENT).		
OS	Rickettsia honei.			
OC	Bacteria: Proteobacteria: alpha subdivision: Rickettsiales;			
CC	Rickettsiaceae: Rickettsiense; Rickettsia.			
OX	NCBI_TaxID=37816;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-TT-118;			
RA	Billings A.N., Yu X.-J., Teel P.D., Walker D.H.;			
RL	Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE-99045882; PubMed-9828442;			
RA	Stenos J., Roux V., Walker D., Raoult D.;			
RL	"Rickettsia honei sp. nov., the aetiological agent of Filanders Island spotted fever in Australia."			
RL	Int. J. Syst. Bacteriol. 48:1399-1404(1998).			
DR	EMBL: AF027124; AAB81846.1; -			
DR	EMBL: AF060706; AAD20231.1; -			
DR	EMBL: AF060704; AAD20230.1; -			
FT	NON TER	1		
SO	SEQUENCE	137 AA:	14167 MW:	75BCLID0D745B428C
Query Match	18.8%:	Score 251.5:	DB 2:	Length 137:
Best Local Similarity	39.4%:	Pred. No. 3.5e-13:		
Matches	52:	Conservative	24:	Mismatches 45; Indels 11; Gaps 4;
OY	120	SNQEVGATGAVGVGAQGLPFKSGGRVSMATG-CAVJLGLGSKTIGSGMDQDK----	174	
DB	10	KNQGTGTLLGGAGGALLSGQFGKGQL-VGVGVALLGAVGGGIGAGMDQDRRLAEL	68	
OY	175	KLNQSLLEKAKAQVRRMRPDTGNSYSVEPVRTYORYNKOEPRQOYCREFOQKAMIAQOK	234	
DB	69	TSQRALLEAPSSSNVEMRRPDMGNNGYTPNKTY-----RNSTGYCRREYTOTVIYIGKQ	123	
OY	235	QEIYGTACPPD	246	
DB	124	OKAYGNACRQPD	135	
RESULT 7				
OJ9K4W8	PRELIMINARY:	PRT:	144 AA:	
AC	OJ9K4W8:			
DT	01-OCT-2000	(TREMBLrel. 15, Created)		
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		

Query Match	18.8%	Score 251.5	DB 2	Length 144
Best Local Similarity	39.4%	Pred. No. 3.7e-13		
Matches 52	Conservative 24	Mismatches 45	Indels 11	Gaps 4
Oy	120	SRQEVGATGAVGVGACGLPKSGSRVSMALG-GAVLGLLGSKTIGSGMDQDK-----I	174	
Db	17	NKQGTGTLGAGGALLGSGFGKRGQL-VGVGVALLGAVLGGOIGAMDSDRLAEL	75	
Oy	175	KINQSLKAKAGVRRMRPDTGNSYSVEPVRYORYNKQERROOYCREFOOKAMTAGOK	234	
Db	76	TSQRALLEAPSSSNVEMRNPDGNGNGYTPNKTY-----RNSTGQYCREYTOTVIYGKQ	130	
Oy	235	QEIYGTACPPD	246	
Db	131	QKSYGNACRPD	142	
RESULT 8				
ID	Q45996	PRELIMINARY	PRT	1546 AA.
AC	Q45996			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE		SCAFOLDING PROTEIN PRECURSOR.		
GN		CIPC.		
OS		Clostridium cellulolyticum.		
OC		Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC		Clostridium.		
OX	NCBI_TaxID=1521;			
RN	[1]			
RP		SEQUENCE FROM N.A.		
RC		STRAIN-ATCC 35319;		
RX		MEDLINE=96218696; PubMed=8636029;		
RA		Pages S., Belaich A., Tardif C., Reverbel-Leroy C., Gaudin C.,		
RA		Belaich J.P.;		
RT		"Interaction between the endoglucanase Cella and the scaffolding		
RT		protein CIPC of the Clostridium cellulolyticum cellulosome.";		
RL		J. Bacteriol. 178:2279-2286(1996).		
RN	[2]			
RP		SEQUENCE FROM N.A.		
RC		STRAIN-ATCC 35319;		
RX		MEDLINE=99173902; PubMed=10074072;		
RA		Pages S., Belaich A., Fierobe H.P., Tardif C., Gaudin C.,		
RA		Belaich J.P.;		
RT		"Sequence analysis of scaffolding protein CIPC and ORFXP, a new		
RT		cohesin-containing protein in Clostridium cellulolyticum: comparison		
RT		of various cohesin domains and subcellular localization of ORFXP.";		
RL		J. Bacteriol. 181:1801-1810(1999).		
RN	[3]			
RP		SEQUENCE FROM N.A.		
RC		STRAIN-ATCC 35319;		
RA		Reverbel-Leroy C., Tardif C., Belaich A., Bernadac A., Gaudin C.,		
RA		Belaich J.P.;		

"Molecular study and overexpression of the Clostridium cellulolyticum celCCE cellulase gene in Escherichia coli.".

RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 35319;

RA Pages S., Belaich A., Reverbel C., Fierobe H.P., Gaudin C., Belaich J.P.;

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RT EMBL: U40345; AAC28899.2; .

DR HSPSP, Q06851; INBC.

DR InterPro: IPR001956; CBD\_3.

DR InterPro: IPR002102; Cohesin.

DR InterPro: IPR003880; Phosphopant\_attach.

DR Pfam: PF00942; CBD\_3; 1.

DR Pfam: PF00963; Cohesin; 8.

DR Prodom: PD001947; CBD\_3; 1.

DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN\_1.

DR Signal.

KW SIGNAL.

FT CHAIN

FT SEQUENCE

1546 AA; 158748 MW; F8651504EC27809F CRC64;

Query Match

Best local Similarity 18.7%; Score 249; DB 2; Length 1546;

Matches 72; Conservative 32; Mismatches 81; Indels 28; Gaps 7;

OY 1 MSVEFYNSKSKQNTSIPITITNTSDSLNDVKKVRYTDTGTCGTCFMCDDHAGAL 60

DB 33 VSVQFNNKSSPANSNSIARFVNTSGPILADLKRIYTTQDADKPLTFMCDHAGYM 92

OY 61 LGNSYVDNTSKYTANFKETASPTSTYDYLDD--PSHMRCLGSSLIITISVFLVCAQ 117

DB 93 SGNNTIDATSKYGTSP-KAVSPAVFNADHYLEVALNSDAGSLPAGSGIEIQTFRANDWS 151

OY 118 NFSROEVAATGAVVGVAGOLFGRKS-----GRVSMATGGAVLGLISKISQSDQDDK 173

DB 152 NFDQSDNMWSTYAA-----GSYMDWKISAFVGGTLAYG-STPDGQNPDPDP 197

OY 174 IKLNOSLEKVKAG-----QVTRMRNPDGTGNSYS 201

DB 198 -TIMPTISAKAGSFADTKITLTPNGNTFNIGIS 229

RESULT 9

053154 PRELIMINARY; PRT; 154 AA.

AC 053154;

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)

DE (CLONE PRB FTISF 1), 5' END CDS (FRAGMENT).

OS Rickettsia sp.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

NCBI\_TaxID=789;

RN [1]

RP SEQUENCE FROM N.A.

RP MEDLINE-93084757; PubMed-1452660.

RA Baird R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;

RT "Characterization and comparison of Australian human spotted fever group rickettsiae.";

RL J. Clin. Microbiol. 30:2896-2902(1992).

DR EMBL: M93931; AAA73386.1; .

FT NON\_TER

FT SEQUENCE

154 AA; 15849 MW; F5C35855EDB439D2 CRC64;

Query Match

Best local Similarity 18.3%; Score 244; DB 2; Length 154;

Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

OY 103 SSIIITISV---FLVGC--AQNFSROEVAATGAVVGVAGOLFGRKSGRVSMATG-GAVL 156

DB 5 SKTIIITATLSMLQACNGPGKMNKGTGTLGAGGALLGSGFGKGGOL-VGVGVGALL 63

OY 157 GGLIGSKISQSDQDDK-----IKLNOSLEKVKAGVTRMRNPDGTGNSYSVPEPRYGRYN 212

DB 64 GAVLGGIGAGMEQDORRLAELTISQRALETAPSGSVNERNPDGNSYGVYTPNKTYRNST 123

OY 213 KQFRQOYCREFOOKAMIAIGOKEIYGTACPPDGR 248

DB 124 GQD-----CRVYTYVYIGSKQKATGNACRQPDQ 154

RESULT 10

09F909 PRELIMINARY; PRT; 151 AA.

AC 09F909;

DT 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE OUTER MEMBRANE PROTEIN (FRAGMENT).

OS Rickettsia helvetica.

OC Bacteria; proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

NCBI\_TaxID=35789;

RN [1]

RP SEQUENCE FROM N.A.

RA Nilsson K., Pahlson C.;

RT "Novel peptide diagnostic reagent and kit for detection of rickettsiosis.";

RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF181036; AAG09427.1; .

FT NON\_TER

FT SEQUENCE

151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match

Best local Similarity 17.9%; Score 239; DB 2; Length 151;

Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

OY 103 SSIIITISV---FLVGC--AQNFSROEVAATGAVVGVAGOLFGRKSGRVSMATG-GAVL 156

DB 5 SKTIIITATLSMLQACNGPGKMNKGTGTLGAGGALLGSGFGKGGOL-VGVGVGALL 63

OY 157 GGLIGSKISQSDQDDK-----IKLNOSLEKVKAGVTRMRNPDGTGNSYSVPEPRYGRYN 212

DB 64 GAVLGGIVAGMEQDORRLAELTISQRALETAPSGSVNERNPDGNSYGVYTPNKTYRNST 119

OY 213 KQFRQOYCREFOOKAMIAIGOKEIYGTACPPDGR 245

DB 120 -RNSTGQYCREYTYVYIGSKQKATGNACRQPD 151

RESULT 11

09F001 PRELIMINARY; PRT; 131 AA.

AC 09F001;

DT 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE 17 KDA PROTEIN (FRAGMENT).

OS Rickettsia sp. California 2.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

NCBI\_TaxID=147259;

RN [1]

RP SEQUENCE FROM N.A.

RP STRAIN-CALIFORNIA 2;

RA Raoult D.;

RT "A new SFG rickettsia isolated from fleas.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RP STRAIN-CALIFORNIA 2;

RA Roux V., Raoult D.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 27, 2002, 11:00:36 ; Search time 9.72712 Seconds  
(without alignments)  
1019.028 Million cell updates/sec

Title: US-09-677-374-6  
Perfect score: 1335  
Sequence: 1 MSVEFYNSKNSAQTNSITPI.....IYGTACPDGMRQVISTEK 256

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	1	CBPA_CLOCL
2	288	21.6	1359	1	17KD_RICPR
3	286	21.4	159	1	17KD_RICJA
4	285	21.3	159	1	17KD_RICCN
5	276.5	20.7	159	1	17KD_RICRY
6	268	20.1	154	1	17KD_RICAU
7	259	19.4	154	1	17KD_RICPA
8	259	19.4	154	1	17KD_RICRH
9	256	19.2	154	1	17KD_RICMO
10	255	19.1	154	1	17KD_RICAM
11	245	18.4	1853	1	CITPA_CLOTM
12	245	18.4	1853	1	CITPA_CLOTM
13	188.5	14.1	879	1	GUNT_CLODM
14	188	14.1	966	1	GUNT_CLOSR
15	162.5	12.2	914	1	GUX2_CLOSR
16	156	11.7	80	1	17KD_RICCA
17	143.5	10.7	1039	1	GUNB_RICCA
18	143.5	10.7	1331	1	MANB_CALSA
19	138.5	10.4	1742	1	GUNA_CALSA
20	121	9.1	499	1	GUN2_BACSU
21	115	8.6	499	1	GUN3_BACSU
22	112.5	8.4	155	1	PCP_YEREN
23	111	8.3	145	1	YCEA_PARELA
24	109.5	8.2	499	1	GUNT_BACSU
25	109	8.2	153	1	SLYB_SALTY
26	104.5	7.8	155	1	SLYB_ECOLI
27	103	7.7	504	1	GUNW_ERMCA
28	101.5	7.6	155	1	PCP_HAEIN
29	101.5	7.6	505	1	GUNV_ERMCA
30	100.5	7.5	700	1	GUNA_PARELA
31	100	7.5	179	1	YCEJ_ECOLI
32	100	7.5	444	1	GUNN_ERMCA
33	98	7.3	922	1	PMPI_CHLPPN

34	97.5	7.3	541	1	NU57_YEAST	P48837	saccharomyc
35	95	7.1	1585	1	YOBO_BACSU	P45931	baecillus su
36	90	6.7	526	1	KICJ_BOVIN	P06394	bos taurus
37	90	6.7	1161	1	DANA_YEAST	P47179	saccharomyc
38	90	6.7	1332	1	KKDO_BACSU	P54334	baecillus su
39	88	6.6	378	1	SLG1_YEAST	P54867	saccharomyc
40	88	6.6	1077	1	HLDS_DROME	Q02308	drosophila
41	87.5	6.6	1061	1	OAR_MYXXA	P38370	myxococcus
42	86.5	6.5	72	1	OSMB_SALTY	P37723	salmonella
43	86.5	6.5	1396	1	VITE_BPR5	P13390	bacterioph
44	86.5	6.5	3770	1	ACVS_EMENTI	P27742	emeritella
45	86	6.4	526	1	VP5_BTIVL	P3476	bluetongue

## ALIGNMENTS

RESULT 1	ID	CBPA_CLOCL	STANDARD:	PRT:	1848 AA.
AC	P38058:	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	Cellulose binding protein A precursor.				
GN	CBPA.				
OS	Clostridium cellulovorans.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;				
OC	Clostridium.				
OX	NCBI_Taxid=1493;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92228810; PubMed=1555642;				
RA	Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;				
RT	"Primary sequence analysis of Clostridium cellulovorans cellulose				
RL	binding protein A.";				
Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).					
CC	-1- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE				
CC	ENZYMES.				
CC	-1- PTM: THE N-TERMINUS IS BLOCKED.				
CC	-1- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib.ch">license@sib.ch</a> ).				
CC	-----				
DR	EMBL: M73817; AAA23218.1; -.				
DR	PIR: A44140; A44140.				
DR	HSSP: Q06851; INBC.				
DR	InterPro: IPR001956; CBD_3.				
DR	InterPro: IPR002102; Cohesin.				
DR	Pfam: PF00942; CBD_3; 1.				
DR	Pfam: PF00963; Cohesin; 9.				
DR	ProDom: PD001947; CBD_3; 1.				
KW	Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.				
FT	SIGNAL				
FT	CHAIN 29 1848				
FT	DOMAIN 29 189				
SO	SEQUENCE 1848 AA; 189152 MW; 85FA6C6F771AF1A CR664;				
QY	Query Match				
QY	Best local similarity 46.8%; Score 487; DB 1; Length 1848;				
QY	Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;				
QY	1 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDVKKRYRYTSRGTOGQGFQFCDHAGAL 60				
QY					
QY	33 MSVEFYNSKNSAQTNSITPIIKITNTSDSLNDVKKRYRYTSRGTOGQGFQFCDHAGAL 92				
QY					
QY	61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPDSHMRCLGSSLIITISVFLVCAQNF 120				

```

Db 93 LGNSYVNTSKVIANFEKETAASPTSTYDYVE-----124
QY 121 ROEVAATGAVGVAGQLFGKSGRYSMAIGAVLGLIGSKIGSMODQDKIKLNSL 180
Db 125 -----FEPASGRATL-----KKQGFIRIQ-----143
QY 181 EKVRAGQVTR--WRNPDTGNSYVEPVRYQRNKRQERROQYERFQOKAMTAGROKEIY 238
Db 144 -----GRITKSDMSWYQTNDYSPDASSSTPPVNP-----KVGYIGGAK--VL 185
QY 239 GTACQPPD 246
Db 186 GTA-PGPD 192

RESULT 2
17KD_RICPR STANDARD; PRT; 159 AA.
ID 17KD_RICPR
AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=69359171; PubMed=2768201;
RA Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
   gene."
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Anderson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
   mitochondria."
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
   anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M28482; AAA26378.1; ALT-SEQ.
DR EMBL: AJ235273; CAI15258.1; -.
DR PIR: D33971; D33971.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16672 MW; A33D04B65EB071 CRC64;
SQ

Query Match 21.6%; Score 288; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 7.2e-17;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
103 SSIIISV---FLVGC--AQNFSROEVAATGAVGVAGQLFGKSGRYSMAIG-GAVL 156

```

```

Db 5 SKMITALASMLQACNGSGMKNKGTGTLGLGAGGALSGQFQGGQL-VGVGVGALL 63
QY 157 GLIGSKIGSMODQDK-----IKLNSLEKVRAGQVTRWRNPDTGNSYVEPVRYQRN 212
Db 64 GAVLGGQIGAGMDEODRRLLETSQRALESAPSGSINEMRPDNGNHGYVPNKTY-----119
QY 213 KOERROQYERFQOKAMTAGROKEIYGTACPOPPDRMOYIS 253
Db 120 -RNSAGYCREYQTGYVIGKQOKRTYGNACROPDGMOMOVN 159

RESULT 3
17KD_RICJA STANDARD; PRT; 159 AA.
ID 17KD_RICJA
AC Q52764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 17 kDa surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH;
RX MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kaiho I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
   specimens by PCR."
RL J. Clin. Microbiol. 33:487-489(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
   anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16515; BAA03965.1; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16554 MW; CDDE7CBBDD0641 CRC64;
SQ

Query Match 21.4%; Score 286; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 1.1e-16;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
103 SSIIISV---FLVGC--AQNFSROEVAATGAVGVAGQLFGKSGRYSMAIG-GAVL 156
Db 5 SKMITALASMLQACNGSGMKNKGTGTLGLGAGGALSGQFQGGQL-VGVGVGALL 63
QY 157 GLIGSKIGSMODQDK-----IKLNSLEKVRAGQVTRWRNPDTGNSYVEPVRYQRN 212
Db 64 GAVLGGQIGAGMDEODRRLLETSQRALESAPSGSINEMRPDNGNHGYVPNKTY-----119
QY 213 KOERROQYERFQOKAMTAGROKEIYGTACPOPPDRMOYIS 253
Db 120 -RNSAGYCREYQTGYVIGKQOKRTYGNACROPDGMOMOVN 159

RESULT 4
17KD_RICCN STANDARD; PRT; 159 AA.
ID 17KD_RICCN
AC P05372;

```

DT	01-NOV-1988 (Rel. 09, Created)
DT	01-AUG-1990 (Rel. 15, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	17 kDa surface antigen precursor.
GN	OMP OR RC1287.
OS	Rickettsia conorii, and
OS	Rickettsia rickettsii.
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC	Rickettsiaceae; Rickettsiense; Rickettsia.
OX	NCBI_TaxID=781, 783;
RA	Anderson B.E., Tzianabos T.;
RA	"Comparative sequence analysis of a genus-common rickettsial antigen
RA	gene".
RA	J. Bacteriol. 171:5199-5201(1989).
RA	[2]
RA	SEQUENCE FROM N.A.
RA	SPECIES=R.conorii, and R.rickettsii;
RA	MEDLINE=89359171; PubMed=2768201;
RA	Anderson B.E., Tzianabos T.;
RA	"Comparative sequence analysis of a genus-common rickettsial antigen
RA	gene".
RA	J. Bacteriol. 171:5199-5201(1989).
RA	[3]
RA	SEQUENCE FROM N.A.
RA	SPECIES=R.rickettsii;
RA	MEDLINE=87222152; PubMed=3108232;
RA	Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA	Fu Z.Y., Bellini W.J.;
RA	"Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
RA	rickettsii".
RA	J. Bacteriol. 169:2385-2390(1987).
RA	[4]
RA	SEQUENCE OF 1-30 FROM N.A.
RA	SPECIES=R.rickettsii;
RA	MEDLINE=89008059; PubMed=3139629;
RA	Anderson B.E., Baumstark B.R., Bellini W.J.;
RA	"Expression of the gene encoding the 17-kilodalton antigen from
RA	Rickettsia rickettsii: transcription and posttranslational
RA	modification".
RA	J. Bacteriol. 170:4493-4500(1988).
RA	[-1 SUBCELLULAR LOCATION: Attached to the outer membrane by a lipiid
RA	anchor (probable).]
RA	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).
CC	-----
CC	EMBL; M28479; AAA26379.1; -
CC	EMBL; M28480; AAA26376.1; -
CC	EMBL; AE008675; AAL03825.1; -
CC	EMBL; M16486; AAA26381.1; -
CC	EMBL; J03571; -; NOT_ANNOTATED_CDS.
CC	PIR; A25972; A25972.
CC	PIR; A25972; A25972.
CC	PIR; A31836; A31836.
CC	PIR; A33971; A33971.
CC	PIR; B33971; B33971.
CC	PROSITE; PS00013; PROKAR_LIPOPROTEIN_1.
CC	Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
CC	SIGNAL
CC	1
CC	19
CC	CHAIN
CC	20 159 17 KDA SURFACE ANTIGEN.
CC	LIPID
CC	20 20 N-ACYL DIGLYCERIDE (PROBABLE).
CC	CONFLICT
CC	146 146 N -> D (IN REF. 3).
CC	FT
CC	CONFLICT
CC	153 153 G -> E (IN REF. 3).
CC	SEQUENCE
CC	159 AA; 16581 MW; 206A2B74FCE1169 CRC64;

[illegible]

RESULT 6  
 17KD\_RICAU STANDARD: PRT: 154 AA.  
 ID 17KD\_RICAU STANDARD: PRT: 154 AA.  
 AC P50928:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia australis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_TaxID=787;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Balid R.W., Ross B., Dwyer B.;  
 RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 anchor (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M74042: AAA26394.1;  
 DR PROSITE: PS00013: PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 19 BY SIMILARITY.  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT NON\_TER 154 154  
 SO SEQUENCE 154 AA; 15967 MW; E3AA83346FAC320 CRC64;  
 Query Match 20.1%; Score 268; DB 1; Length 154;  
 Best Local Similarity 37.8%; Pred. No. 3e-15;  
 Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;  
 OY 103 SLLIIIV---FLVGCAC--NFSRQEVGAATGAVVGAGQLFGKSGRYSMAIG-GAVL 156  
 DB 5 SKIMIALAASMLQACNPGMKNKOGTGTLLGAGALLSQKKGQL-VGVGALL 63  
 OY 157 GGLISKTIQGSMDQDK-----IKLNQSLKVKAGQVTRMRNPTGNSYSVEPVRTQRYN 212  
 DB 64 GAVLGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNDNGYVTPNKTYRNSN 123  
 OY 213 KQRRQOYCREFOQKAMTAGOKOELIYGTACPODGR 248  
 DB 124 -----GQYCREYOTVTVIGKQOKAYGNACLPDQ 154  
 RESULT 7  
 17KD\_RICPA STANDARD: PRT: 154 AA.  
 ID 17KD\_RICPA STANDARD: PRT: 154 AA.  
 AC P50930:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia parkeri.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_TaxID=35792;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Pretzman C.I., Stochard D.R., Ralph D., Clark J.B., Fuerst P.A.;  
 RC

Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 anchor (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U17008: AAA82040.1;  
 DR PROSITE: PS00013: PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 19 BY SIMILARITY.  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT NON\_TER 154 154  
 SO SEQUENCE 154 AA; 15897 MW; 5D06F45F9BDEEC CRC64;  
 Query Match 19.4%; Score 259; DB 1; Length 154;  
 Best Local Similarity 36.5%; Pred. No. 1.7e-14;  
 Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;  
 OY 103 SLLIIIV---FLVGC--AQNFSRQEVGAATGAVVGAGQLFGKSGRYSMAIG-GAVL 156  
 DB 5 SKIMIALAASMLQACNPGMKNKOGTGTLLGAGALLSQKKGQL-VGVGALL 63  
 OY 157 GGLISKTIQGSMDQDK-----IKLNQSLKVKAGQVTRMRNPTGNSYSVEPVRTQRYN 212  
 DB 64 GAVLGQIGAGMDEDRRLAELTSGRALETAPSGSNVEMRNDNGYVTPNKTYRNSN 119  
 OY 213 KQRRQOYCREFOQKAMTAGOKOELIYGTACPODGR 248  
 DB 120 -RNSTGQYCREYOTVTVIGKQOKAYGNACLPDQ 154  
 RESULT 8  
 17KD\_RICRH STANDARD: PRT: 154 AA.  
 ID 17KD\_RICRH STANDARD: PRT: 154 AA.  
 AC P50931:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE 17 kDa surface antigen precursor (Fragment).  
 GN OMP.  
 OS Rickettsia rhipicephali.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_TaxID=31992;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid  
 anchor (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U11020: AAB0706.1;  
 DR PROSITE: PS00013: PROKAR\_LIPOPROTEIN; 1.  
 KW Outer membrane; Lipoprotein; Antigen; Signal.  
 FT SIGNAL 19 BY SIMILARITY.  
 FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.  
 FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT NON\_TER 154 154



SEQUENCE 154 AA: 15895 MW: 0CF85AD5D96DFE6B CRC64:

Query Match 19.4%; Score 259; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 1.7e-14; Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 103 SLLIISV---FLVGC--AQNFSSROEVGAATGAVVGGVAGOLPFKSGSRVSMATG-GAVL 156

DB 5 SKIMITIALASMLQACNGPGGMKQGTGTLGAGGALLGSQFQKGGQL-VGVGVGALL 63

QY 157 GGLISKRIGSMDQDK---IKLNQSLKVKAGQVTRMRNPDGNSVSEVPVRYORYN 212

DB 64 GAVLGCGIGAGMDQDRRLAELTSGRALETAPSGSNVEMRNPDMNGNYITPNKTY----- 119

QY 213 KOERROOYCFRFOOKAMIAQOKETITGTACPOPDR 248

DB 120 -RNSTGOYCREYTOTVYIGGKQKAYGNACLOPDGQ 154

RESULT 9

17KD\_RICMO STANDARD: PRT; 154 AA.

AC P50929;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 Kda surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia montana.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_TaxID=33991;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OHIO 83-441;

RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid

CC anchor (Probable).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: U11017; AAB07705.1; -

DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).

FT NON\_TER 154 154

SO SEQUENCE 154 AA: 15881 MW: A09C53B8769E31DA CRC64;

Query Match

Best Local Similarity

Matches 57; Conservative

QY 103 SLLIISV---FLVGC--AQNFSSROEVGAATGAVVGGVAGOLPFKSGSRVSMATG-GAVL 156

DB 5 SKIMITIALASMLQACNGPGGMKQGTGTLGAGGALLGSQFQKGGQL-VGVGVGALL 63

QY 157 GGLISKRIGSMDQDK---IKLNQSLKVKAGQVTRMRNPDGNSVSEVPVRYORYN 212

DB 64 GAVLGCGIGAGMDQDRRLAELTSGRALETAPSGSNVEMRNPDMNGNYITPNKTY----- 119

QY 213 KOERROOYCFRFOOKAMIAQOKETITGTACPOPDR 248

DB 120 -RNSTGOYCREYTOTVYIGGKQKAYGNACLOPDGQ 154

RESULT 11

CIPB\_CLOTM STANDARD: PRT; 772 AA.

AC 001866;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SL/SL)

DE Cellulose integrating protein B (Fragment).

GN CIPB.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium.

OX NCBI\_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS.

```

RX MEDLINE-93146373; PubMed-1490597;
RA Poole D.M., Morag E., Lamed R., Bayer E.A., Hazlewood G.P.,
RA Gilbert H.J.;
RT "Identification of the cellulose-binding domain of the cellulosome
RT strain S1 from Clostridium thermocellum YS.";
RL FEBS Microbiol. Lett. 78:181-186(1992).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
CC PRESENT IN CATALYTIC SUBUNIT OF THE CELLULOSE.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68233; CAA48312.1; -.
DR HSSP: 006851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cohesin.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00963; Cohesin; 3.
DR Pfam: PF00404; Dockerin_1; 2.
DR ProDom: PD001947; CBD_3; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
FT DOMAIN 1 1 607 3 X 147 AA APPROXIMATE REPEATS.
FT REPEAT <1 80 1 (PARTIAL).
FT REPEAT 94 240 2.
FT REPEAT 462 607 3.
FT DOMAIN 81 93 PRO/THR-RICH.
FT DOMAIN 241 272 PRO/THR-RICH.
FT DOMAIN 273 439 CELLULOSE-BINDING.
FT DOMAIN 440 461 PRO/THR-RICH.
FT DOMAIN 710 766 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 710 733 2-1.
FT REPEAT 743 766 2-2.
SO SEQUENCE 772 AA; 82491 MW; BBF06DE5E094FE10 CRC64;

Query Match 18.4%; Score 245; DB 1; Length 772;
Best Local Similarity 51.5%; Pred No. 1.5e-12;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSVEEYNSKSAQNTSITPIKINTSDSDINLMDVKVRYIYSDGQGTFCMDHAGAL 60
DB 281 LKVEEYNSPSTNSINPQKVTNIGSSAIDLSKLTLYRYIYDGGKDTFMDHA-AI 339
QY 61 LCN--SYVDNISKYANFVKETASPTSYDYIDPSHMRGLQ 101
DB 340 IGSNGSYNGITSNVKGTFVKMSSS-TNNADYILEISTFGTLE 381

RESULT 12
CIPA_CLOTM STANDARD: PRT; 1853 AA.
AC 006851;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cellulosomal scaffolding protein A precursor (cellulosomal
DE glycoprotein S1/SL) (Cellulose integrating protein A) (Cohesin).
OS CIPA.
OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
RC STRAIN-ATCC 27405;
RA MEDLINE-9320508; PubMed-8316083;
RA Gerngross U.T., Romaniec M.P.M., Kobayashi T., Haskisson N.S.,
RA Demain A.L.;
RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the
RT cellulosomal SL-protein reveals an unusual degree of internal
RT homology.";
RL Mol. Microbiol. 8:325-334(1993).
RN [2]
RP SEQUENCE OF 1820-1853 FROM N.A.
RX MEDLINE-9320931; PubMed-8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE-97238934; PubMed-9083107;
RA Shimon I.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolov F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly.";
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE-97076134; PubMed-8918451;
RA Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steltz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose.";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE-98022914; PubMed-9402065;
RA Tavares G.A., Beguin P., Alzari P.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution.";
RL J. Mol. Biol. 273:701-713(1997).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
CC COMPONENTS OF THE CELLULOSE.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08665; -. NOT_ANNOTATED_CDS.
DR EMBL: X67506; CAA47840.1; -.
DR PIR: S36859; S36859.
DR PDB: 1ANU; 23-JUL-97.
DR PDB: 1AOK; 08-JUL-98.
DR PDB: INBC; 26-SEP-97.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002102; Cohesin.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.

```

DR Pfam: PF00942; CBD\_3; 1.  
 DR Pfam: PF00963; Cohesin; 9.  
 DR Pfam: PF00404; Dockerin\_1; 2.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
 DR PROSITE: PS00448; CLOS\_CELLOSOME\_RPT; 2.  
 KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;  
 3D-structure.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1853  
 FT DOMAIN 364 522  
 FT DOMAIN 54 1694  
 FT REPEAT 29 182  
 FT REPEAT 183 322  
 FT REPEAT 560 704  
 FT REPEAT 724 866  
 FT REPEAT 889 1031  
 FT REPEAT 1054 1196  
 FT REPEAT 1219 1361  
 FT REPEAT 1384 1526  
 FT REPEAT 1548 1690  
 FT DOMAIN 1791 1846  
 FT REPEAT 1791 1814  
 FT REPEAT 1824 1847  
 FT REPEAT 1615 1615  
 FT CONFLICT 1615 1615  
 SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 18.48; Score 245; DB 1; Length 1853;  
 Best Local Similarity 51.58; Pred. No. 4; Ie-12;  
 Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSVEFYNSKSAQTNSITPIKIKINTSDSLNDVKVRYRYTSDGTGOTGFWCDHAGAL 60  
 Db 369 LKVEFYNSNPDTNTSINPQKVTNTGSSAIDLSKLTIRYTYVDGQKQDTEWCDHA-AI 427  
 61 LGN--SYVDNTSKVTANFKETASPTSTYDTYLDPSHMRGCLQ 101  
 Db 428 IGSNGSYNGITSNKGTFVKMSS--TNNADYLETLSFTGYLE 469

RESULT 13  
 GUN2\_CLOTRM STANDARD; PRT; 879 AA.  
 AC 002934;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Endoglucanase I precursor (EC 3.2.1.4) (Egri) (Endo-1,4-beta-glucanase)  
 DE (Cellulase I).  
 CN CELL.  
 OS Clostridium thermocellum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_Taxid=1515;  
 RN [1].  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.  
 RC STRAIN=NCIB 10662; PubMed=8436949;  
 RA Haelewood G.P., Davidson K., Laurie J.I., Hushisson N.S.,  
 RA Gilbert H.J.;  
 RT "Gene sequence and properties of Cell, a family E endoglucanase from  
 Clostridium thermocellum".  
 RL J. Gen. Microbiol. 139:307-316(1993).  
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
 GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
 GLUCANS. PRINCIPALLY ACTIVE AGAINST BARLEY BETA-GLUCAN.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose.  
 CC -1- PATHWAY: CELLULOSE DEGRADATION.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL  
 HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I04735; AAA20892.1; -.  
 DR PIR: A47704; A47704.  
 DR HSSP: P26221; IFE4.  
 DR InterPro: IPR001956; CBD\_3.  
 DR InterPro: IPR001701; Glyco\_hydro\_9.  
 DR Pfam: PF00942; CBD\_3; 2.  
 DR Pfam: PF00759; Glyco\_hydro\_9; 1.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR PROSITE: PS00592; GLYCOSYL\_HYDROL\_P9\_1; 1.  
 DR PROSITE: PS00698; GLYCOSYL\_HYDROL\_P9\_2; 1.  
 KW Cellulose degradation; Hydrolyase; Glycosidase; Signal.  
 FT SIGNAL 1 55  
 FT CHAIN 56 879  
 FT DOMAIN 56 518  
 FT DOMAIN 729 879  
 FT ACT\_SITE 448 448  
 FT ACT\_SITE 486 486  
 FT ACT\_SITE 495 495  
 FT CONFLICT 495 495  
 SQ SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 14.18; Score 188.5; DB 1; Length 879;  
 Best Local Similarity 28.48; Pred. No. 7; Je-08;  
 Matches 59; Conservative 20; Mismatches 48; Indels 81; Gaps 9;

QY 3 VEFYNSKSAQTNSITPIKIKINTSDSLNDVKVRYRYTSDGTGOTGFWCDHAGALLG 62  
 Db 742 LQYANGNAGATSNISNPFKIKINNGTKAINDVKIRIRYTKEGASQNFQCDHSSA--G 799  
 63 NSYDNTSKVTANFKETASPTSTYDTYLDPSHMRGCLQSSLIIVSVLGCANQNSRQ 122  
 Db 800 N-----SNVTGNF--NLSPEKAGADT-----CL----- 821

QY 123 EVGAATGAVVGAVGQLEFGKSGRYSMAGAVLGLGSKIGSMQDQKIKINQSLER 182  
 Db 822 EVG-----FGSGAG--TLDPGGSV-----EVQIRFSKED 848  
 QY 183 VKAGQVTRMRPDPDNGNSVEPRTYR 210  
 Db 849 -----WSNYNOSNDYSFKQACLRQR 868

RESULT 14  
 GUN2\_CLOTR STANDARD; PRT; 986 AA.  
 AC P23659;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Endoglucanase Z precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Thermocactive cellulase) (Avicelase I).  
 CN CELL.  
 OS Clostridium stercorarium.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_Taxid=1510;  
 RN [1].  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.  
 RC STRAIN=NCIB 11745;  
 RA MEDLINE=91066838; PubMed=2250652;  
 RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,  
 RA Bronnenmeier K., Staudenbauer W.L.;  
 RT "Sequence analysis of the Clostridium stercorarium celZ gene encoding  
 a thermocactive cellulase (Avicelase I): Identification of catalytic  
 RT and cellulose-binding domains".  
 RL Mol. Gen. Genet. 223:258-267(1990).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

linkages in cellulose.  
 CC -1- SIMILARITY: CONTAINS MULTIPLE CELLULOSE-BINDING SITES.  
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X55299; CAA39010.1; ALT\_SEQ.  
 DR PIR: S12021; S12021.  
 DR HSSP: P26221; 1TF4.  
 DR InterPro: IPR001956; Glyco\_hydro\_9.  
 DR InterPro: IPR001701; Glyco\_hydro\_9.  
 DR Pfam: PF00942; CBD\_3; 2.  
 DR Pfam: PF00759; Glyco\_hydro\_9; 1.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR PROSITE: PS00592; GLYCOSYL\_HYDROL\_F9\_1; 1.  
 DR PROSITE: PS00598; GLYCOSYL\_HYDROL\_F9\_2; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 KW SIGNAL  
 FT CHAIN 1 26 986  
 FT REPEAT 498 641  
 FT REPEAT 651 738  
 FT REPEAT 744 831  
 FT REPEAT 854 986  
 FT DOMAIN 835 986  
 FT ACT\_SITE 400 400  
 FT ACT\_SITE 438 438  
 FT ACT\_SITE 447 447  
 SQ SEQUENCE 986 AA: 109512 MW: 180280982923690 CRC64;  
 Query Match 14.1%; Score 188; DB 1; Length 986;  
 Best Local Similarity 27.6%; Pred. No. 9.2e-08;  
 Matches 47; Conservative 29; Mismatches 58; Indels 36; Gaps 6;  
 QY 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNDVRYRYTSDGTGQTFWCHAGAL 60  
 Db 840 IQIDFMNGNSDKTNGIMPRKRLTCTPIRLSDVKIRYYTIDEDKQDNWCWSSV- 898  
 QY 61 LGNSYVDNTSKVTANFYKETASPTSTYDTYLDSSHMGCC-LOGSSLIITSVFLVCAON 118  
 Db 899 -----GSSNITGTFFVK-MAEPKEGADYYLETGFTDGAGTLOPNOSIEVQ-----NR 943  
 QY 119 FSR-----QEVGATGAVGVGAGVAGQLFGKSGRVSMAIGAVLGL 159  
 Db 944 FSKADMTDYIOTNDYSTSTNTSYG-----SNDRIITYISGVAVSGI 984  
 RESULT 15  
 GUX2\_CLOSR STANDARD: PRT: 914 AA.  
 ID GUX2\_CLOSR  
 AC P50900;  
 DT 01-OCT-1996 (rel. 34, Created)  
 DT 01-OCT-1996 (rel. 34, Last sequence update)  
 DT 01-NOV-1997 (rel. 35, Last annotation update)  
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocelllobiohydrolase II)  
 DE (1,4-beta-cellobiohydrolase II) (Avicelase II).  
 GN CELV.  
 OS Clostridium stercorarium.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCB1\_TaxID=1510;  
 RN SROUENCE FROM N.A.  
 RA STRAIN-NCIB 11754;  
 RA Staudenbauer W.L.;  
 RA Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.

CHARACTERIZATION.  
 RN STRAIN-NCIB 11754;  
 RC MEDLINE-91364686; PubMed-1909625; Staudenbauer W.L.;  
 RA Bromenneier K., Ruecknagel K.P., Staudenbauer W.L.;  
 RT "Purification and properties of a novel type of  
 RT exo-1,4-beta-glucanase (avicelase II) from the cellulolytic  
 RT thermophile Clostridium stercorarium";  
 RL Eur. J. Biochem. 200;379-385(1991).  
 CC -1- CATALYTIC ACTIVITY: Hydrolase of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellotetraose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z69359; CAA93280.1; -  
 DR HSSP: O06851; INBC.  
 DR InterPro: IPR001956; CBD\_3.  
 DR InterPro: IPR00556; Glyco\_hydro\_48.  
 DR Pfam: PF00942; CBD\_3; 1.  
 DR Pfam: PF00844; GLHYDRLASE48.  
 DR PRINTS: PR00844; GLHYDRLASE48.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR ProDom: PD011903; Glyco\_hydro\_48; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 KW SIGNAL  
 FT CHAIN 1 34 914  
 FT SIGNAL 34 914  
 SQ SEQUENCE 914 AA: 103020 MW: D0DB6017D6DFF82C CRC64;  
 Query Match 12.2%; Score 162.5; DB 1; Length 914;  
 Best Local Similarity 37.8%; Pred. No. 1e-05;  
 Matches 34; Conservative 17; Mismatches 30; Indels 9; Gaps 2;  
 QY 3 VEFYNSKSAQNTSTPIIKITNTSDSLNDVRYRYTSDGTGQTFWCHAGALLG 62  
 Db 769 IQSFNATQEIINSIMPRFRIYNSGNTSIPLSVKKLRYTYVDGKPKQNFQCDMSI--- 825  
 QY 63 NSYVDNTSKVTANFYKETASPTSTYDTYLD 92  
 Db 826 -----GSSNITGTFFVKMDGATGGA-DYYLE 849  
 Search completed: October 27, 2002, 11:10:53  
 Job time : 11.7271 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:05:56 ; Search time 16.3592 Seconds  
(without alignments)  
1503.668 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335

Sequence: 1 MSVEFYNSNKSAGTNSITPI.....ITGTACPDGDMQVISTEK 256

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR\_71:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	2 A44140	cellulose-binding
2	288	21.6	159	2 D33971	Rickettsial common
3	285	21.3	159	2 B33971	Rickettsial common
4	285	21.3	159	2 A33971	Rickettsial common
5	285	21.3	159	2 G97860	17K surface antigen
6	276.5	20.7	159	2 C33971	Rickettsial common
7	276	20.7	159	2 A25972	17K antigen precursor
8	249	18.7	586	2 PC6006	scaffoldin proteol
9	245	18.4	1854	2 S36859	cipa protein - C10
10	233.5	17.5	1162	2 T30433	scaffoldin proteol
11	223.5	16.7	1483	2 G97012	probably cellulose
12	188.5	14.1	879	2 A47704	endoglucanase I (E
13	188	14.1	986	2 S12021	thermoactive cellu
14	151.5	11.3	1779	2 T31085	xylanase - Caldice
15	147	11.0	915	2 A43802	cellulase (EC 3.2.
16	143.5	10.7	1039	2 S02711	cellulase (EC 3.2.
17	143.5	10.7	1331	2 A48954	mannan endo-1,4-be
18	142	10.6	1711	2 T31337	1,4-beta-glucanase
19	138.5	10.4	1742	2 T17120	cellulase (EC 3.2.
20	127.5	9.6	182	2 D83169	conserved hypotet
21	122	9.1	257	2 F87413	hypothetical prote
22	121	9.1	508	2 G69593	cellulase (EC 3.2.
23	120.5	9.0	131	2 AI3418	17K surface antigen
24	115	8.6	499	2 JN0111	cellulase (EC 3.2.
25	113.5	8.5	499	2 A27198	cellulase (EC 3.2.
26	112.5	8.4	155	2 S23787	outer membrane lip
27	111.5	8.4	154	2 B83514	conserved hypotet
28	111	8.3	145	2 A41897	cellulase homology
29	109.5	8.2	508	2 A26874	cellulase (EC 3.2.

30	109	8.2	155	2 A10693	outer membrane lip
31	109	8.2	179	2 AE0644	probable secreted
32	109	8.2	232	2 G87629	hypothetical prote
33	109	8.2	257	2 B82837	conserved hypotet
34	107.5	8.1	142	2 AD2696	lipa protein (limpo
35	107.5	8.1	155	2 AF0289	probable lipoprote
36	107	8.0	125	2 D97438	lipa protein (limpo
37	107	8.0	486	2 I40548	bifunctional cellu
38	106.5	8.0	155	2 AG0443	outer membrane lip
39	104.5	7.8	155	2 C64921	outer membrane lip
40	104.5	7.8	155	2 F90922	probable outer mem
41	104.5	7.8	155	2 B85771	probable outer mem
42	104	7.8	139	2 S58234	lipa protein - Rhl
43	103	7.7	232	2 AD3350	outer membrane pro
44	103	7.7	504	2 S54744	cellulase (EC 3.2.
45	101.5	7.6	155	2 I64130	PAL cross-reacting

#### ALIGNMENTS

##### RESULT 1

A44140

cellulose-binding protein A precursor - Clostridium cellulovorans

C:Species: Clostridium cellulovorans

C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 15-Oct-1999

C:Accession: A44140

R:Shoseyov, O.; Takagi, M.; Goldstein, M.A.; Doi, R.H.

Proc. Natl. Acad. Sci. U.S.A. 89, 3483-3487, 1992

A:Title: Primary sequence analysis of Clostridium cellulovorans cellulose binding pro

A:Reference number: A44140; MUID:92228810

A:Accession: A44140

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1-1848 <SHO>

A:Cross-references: GB:M73817; NID:g144748; PIDN:AAA23218.1; PID:g144749

C:Genetics:

A:Gene: cbpa

Query Match 36.5%; Score 487; DB 2; Length 1848;

Best Local Similarity 46.8%; Pred. No. 2.8e-32;

Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

QY	1	MSVEFYNSNKSAGTNSITPIIKITNSDSDLNLDVAVRRYRYSDDGQGTFCMDHAGAL	60
DB	33	MSVEFYNSNKSAGTNSITPIIKITNSDSDLNLDVAVRRYRYSDDGQGTFCMDHAGAL	92
QY	61	LGNSYVDNTSKYANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVGCAQNF	120
DB	93	LGNSYVDNTSKYANFVKETASPTSTYDTYLDPSHMRGCLQGSLLIISVFLVGCAQNF	124
QY	121	ROEVAATGAAGVAGVAGQLFGKSGRVSMAIGAVLGLIGSKIGSGMDQDKIKLNQSL	180
DB	125	-----FGFASGRATL-----KKGQITIQ-----	143
QY	181	EKKVAGQVTR--WRNPDTGSSVSEPRVRYORNYKQRRQOYCREPQKAMINGQKEIY	238
DB	144	-----GRITKSDMSNVTQTDYSPDASSRPVYVP-----KVYIGIGAK--VL	185
QY	239	GTACPOP 246	
DB	186	GTACPOP 192	

##### RESULT 2

D33971

Rickettsial common antigen precursor - Rickettsia prowazekii

N:Alternate names: 17kD surface antigen; outer membrane protein (omp); RP833

C:Species: Rickettsia prowazekii

C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 03-Nov-2000

C:Accession: D33971; B71645

R:Anderson, B.E.; Tzianabos, T.

J. Bacteriol. 171, 5199-5201, 1989



RESULT 8  
PC6006  
scaffolding protein C1pC precursor - Clostridium cellulolyticum (fragment)  
C:Species: Clostridium cellulolyticum  
C:Date: 17-Aug-1990 #sequence\_revision 16-Aug-1996 #text\_change 13-Nov-1998  
C:Accession: PC6006  
J. Pagers, S.; Belatch, A.; Tardif, C.; Reverdel-Perry, C.; Gaudin, C.; Belatch, J.P.  
J. Bacteriol. 178, 2279-2286, 1996

RESULT 9

536859

c1pA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL

C:Species: Clostridium thermocellum

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000

C:Accession: S36859; S33527; S25767; S28659; T18261

R:Gerrgros, U.T.; Demain, A.L.

submitted to the EMBL Data Library, January 1993

A:Reference number: S36859

A:Status: preliminary

A:Accession: S36859

A:Molecule type: DNA

A:Residues: 1-1854 <GER>

A:Cross-references: EMBL:L08665

R:Gerrgros, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.

Mol. Microbiol. 8, 325-334, 1993

A:Title: Sequencing of a Clostridium thermocellum gene (c1pA) encoding the cellulosome

A:Reference number: S33527; MUID:93302508

A:Accession: S33527

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1615,1617-1854 <GE2>

A:Cross-references: EMBL:L08665

R:Fujino, T.; Beguin, P.; Aubert, J.P.

FEMS Microbiol. Lett. 94, 165-170, 1992

A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides tha

A:Reference number: S25767

A:Accession: S25767

A:Molecule type: DNA

A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FUG>

A:Cross-references: EMBL:X67406

R:Bequin, P.

submitted to the EMBL Data Library, August 1992

A:Reference number: S28659

A:Accession: S28659





Db 849 -----MSNYNSNDYSEFKOACLROR 868

## RESULT 13

S12021

thermoactive cellulase - Clostridium stercorarium  
N:Alternate names: avicelase I

C:Species: Clostridium stercorarium

C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997

C:Accession: S12021

R:Authors, S.: Ruecknagel, K.P.; Schwarz, W.H.; Kratzsch, P.; Bronnenmeier, K.; Staudenb

Mol. Gen. Genet. 223, 258-267, 1990

A:Title: Sequence analysis of the Clostridium stercorarium celz gene encoding a thermoac

A:Reference number: S12021; MUID:9106838

A:Accession: S12021

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-986 <JAU>

Query Match 14.1%; Score 188; DB 2; Length 986;  
Best Local Similarity 27.6%; Pred. No. 1.4e-07;

Matches 47; Conservative 29; Mismatches 58; Indels 36; Gaps 6;

QY 1 MSVEFYNSNKAQNTSITPIIKITNTSDSLNDVKVRYVYSDGTQGGTFWCDAHAL 60

DB 840 IQIQMFNGNTSDKTNGLMPRYRLNTGTPIRLSDVKIRYYITIDGEMQFMCWMSV- 898

QY 61 LGNSVYDNTSKVTANFVKETASPTSTYDYLDPHMRGC--LQSSLLIISVFLVCAON 118

DB 899 -----GSNNITGTFFVK-MAEPKEGADYYLETGFTDAGYLQPNQSIQVQ-----NR 943

QY 119 FSR-----OEVGATGAVGVGAGOLF-GKSGRVSMAIGAVLGL 159

DB 944 FSKADMWDIYDTNDYSEFTNTSYG-----SNDKITVYISVLVSGI 984

## RESULT 14

T31085

xylinase - Caldicellulosiruptor sp.

C:Species: Caldicellulosiruptor sp.

C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL Data Library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.

A:Reference number: Z20972

A:Accession: T31085

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1779 <MOR>

A:CROSS-references: EMBL:AF036924; NID:q2760905; PID:q2760909; PIDN:AB95326.1

C:Genetics:

A>Note: xynC

Query Match 11.3%; Score 151.5; DB 2; Length 1779;

Best Local Similarity 33.6%; Pred. No. 0.00032;

Matches 51; Conservative 15; Mismatches 35; Indels 51; Gaps 6;

QY 1 MSVEFYNSNKAQNTSITPIIKITNTSDSLNDVKVRYVYSDGTQGGTFWCDAH--G 58

DB 1114 LKVLKNNETASSTIRPMFKIVNGSSVDSLVRKIRIYWTVDGKPGQAVCDMAQIG 1173

QY 59 ALLGNSYVDNTSKVTANFVKETASPTSTYDYLDPHMRGCLQSSLLIISVFLVCAON 118

DB 1174 A-----SNVTFNFKVLT-SGVSGADYLL----- 1195

QY 119 FSRQEVGAATGAVGVAGOLF-GKSGGRVSM 149

DB 1196 -----EVGFSSG-----AGQLQPKGKDTGDIQV 1217

## RESULT 15

A43802

cellulase (EC 3.2.1.4) / cellulose 1,4-beta-D-glucosidase (EC 3.2.1.91) - Caldocellu

N:Alternate names: endo-1,4-beta-D-glucanase

C:Species: Caldocellum saccharolyticum 30-Sep-1993 #text\_change 10-Jul-1998

C>Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993

C:Accession: A43802

R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P

Appl. Environ. Microbiol. 56, 3117-3124, 1990

A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophil

A:Reference number: A43802; MUID:91136262

A:Accession: A43802

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: DNA

A:Residues: 1-915 <SAU>

A:CROSS-references: EMBL:X13602

C:Function: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F;20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 11.0%; Score 147; DB 2; Length 915;  
Best Local Similarity 23.2%; Pred. No. 0.00034;

Matches 60; Conservative 35; Mismatches 64; Indels 100; Gaps 11;

QY 1 MSVEFYNSNKAQNTSITPIIKITNTSDSLNDVKVRYVYSDGTQGGTFWCDAH--G 58

DB 351 IKVLYANKETNSTNTIRPMFKIVNCGSSSIDLSRYITRTYVDGERAOSAVSDMAQIG 410

QY 59 ALLGNSYVDNTSKVTANFVKETASPTSTYDYLDPHMRGCLQSSLLIISVFLVCAON 118

DB 411 A-----SNVTFNFKVLT-SSSVSGADYLL-----QIRFN 432

QY 119 FSRQEVGAATGAVGVAGOLF-GKSGRVSMAIGAVLGLGSLGSKIGQSMDDQDKIKLN 177

DB 433 -----EIGFKSG-----AGQLQPKGKDTGEI----- 457

QY 178 QSLKRVKAGVTRMRNDPTGNSYSEVPEYRTQRYNKOERRQYCRFQOKAMINGOKOI 237

DB 458 KS-----DMSNTNCGNDMSW--LQSMSTYGENEKYATYI---DGVLVWGQE--- 498

QY 238 YGTACPOPDGRWQVISTEK 256

DB 499 -----PSTNDMLVYSGNK 512

Search completed: October 27, 2002, 11:11:39  
Job time : 19.3592 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 11:08:02 : Search time 13.2642 Seconds  
(without alignments)  
471.414 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335  
Sequence: 1 MSVEFYNSNKSQNTSITPI.....ITGACPDGRNQVISTEK 256

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/p/odata/1/1aa/5A.COMB.pep:\*

2: /cgn2\_6/p/odata/1/1aa/5B.COMB.pep:\*

3: /cgn2\_6/p/odata/1/1aa/6A.COMB.pep:\*

4: /cgn2\_6/p/odata/1/1aa/6B.COMB.pep:\*

5: /cgn2\_6/p/odata/1/1aa/6C.COMB.pep:\*

6: /cgn2\_6/p/odata/1/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	480.5	36.0	382	4 US-09-277-716-22	Sequence 22, Appl
2	477	35.7	162	1 US-08-048-164A-2	Sequence 2, Appl
3	477	35.7	162	1 US-08-460-462-2	Sequence 2, Appl
4	477	35.7	162	1 US-08-460-457-2	Sequence 2, Appl
5	477	35.7	162	1 US-08-460-458-2	Sequence 2, Appl
6	477	35.7	162	1 US-08-460-455-2	Sequence 2, Appl
7	477	35.7	162	2 US-08-330-394A-2	Sequence 2, Appl
8	477	35.7	163	3 US-09-006-636-7	Sequence 7, Appl
9	477	35.7	163	4 US-09-006-632-7	Sequence 29, Appl
10	472	35.4	154	2 US-08-330-394A-29	Sequence 22, Appl
11	472	35.4	156	2 US-08-330-394A-22	Sequence 22, Appl
12	472	35.4	167	5 PCT-US95-13813-9	Sequence 9, Appl
13	245	18.4	493	4 US-09-198-956-10	Sequence 10, Appl
14	245	18.4	493	4 US-09-198-955A-12	Sequence 12, Appl
15	149.5	11.2	1426	4 US-09-136-574A-43	Sequence 43, Appl
16	148.5	11.1	616	4 US-09-136-574A-47	Sequence 47, Appl
17	148.5	11.1	1751	4 US-09-136-574A-44	Sequence 44, Appl
18	122.5	9.2	551	2 US-09-033-537A-1	Sequence 1, Appl
19	100.5	7.5	700	2 US-07-862-588B-2	Sequence 12, Appl
20	85	6.4	1021	1 US-07-910-760-12	Sequence 12, Appl
21	85	6.4	1021	1 US-08-440-519-12	Sequence 12, Appl
22	85	6.4	1021	1 US-08-440-549-12	Sequence 12, Appl
23	83	6.2	382	4 US-08-444-818-68	Sequence 68, Appl
24	83	6.2	460	4 US-08-444-818-20	Sequence 58, Appl
25	83	6.2	512	4 US-08-867-611-58	Sequence 47, Appl
26	83	6.2	592	4 US-08-867-611-47	Sequence 47, Appl
27	83	6.2	594	4 US-08-867-611-48	Sequence 48, Appl

28	83	6.2	597	4 US-08-867-611-16	Sequence 16, Appl
29	83	6.2	597	5 PCT-US92-06965A-21	Sequence 21, Appl
30	83	6.2	599	4 US-08-867-611-18	Sequence 18, Appl
31	83	6.2	599	5 PCT-US92-06965A-23	Sequence 23, Appl
32	83	6.2	613	4 US-08-867-611-49	Sequence 49, Appl
33	83	6.2	739	4 US-08-444-818-148	Sequence 148, Appl
34	83	6.2	781	4 US-08-867-611-4	Sequence 4, Appl
35	83	6.2	781	5 PCT-US92-06965A-9	Sequence 9, Appl
36	83	6.2	859	1 US-08-444-818-30	Sequence 30, Appl
37	83	6.2	970	1 US-08-375-709-7	Sequence 7, Appl
38	83	6.2	970	1 US-08-752-929-7	Sequence 7, Appl
39	83	6.2	970	4 US-09-090-793-5	Sequence 5, Appl
40	83	6.2	971	4 US-08-867-611-52	Sequence 52, Appl
41	83	6.2	973	4 US-08-867-611-53	Sequence 53, Appl
42	83	6.2	992	4 US-08-867-611-54	Sequence 54, Appl
43	83	6.2	1786	4 US-08-444-818-54	Sequence 54, Appl
44	83	6.2	2261	4 US-08-444-818-66	Sequence 66, Appl
45	83	6.2	2772	4 US-08-444-818-89	Sequence 89, Appl

## ALIGNMENTS

RESULT 1  
US-09-277-716-22  
Sequence 22, Application US/09277716A  
Patent No. 6232107  
GENERAL INFORMATION:  
APPLICANT: Bryan, Bruce  
APPLICANT: Szent-Gyorgyi, Christopher  
APPLICANT: PROUME, LTD.  
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE  
CURRENT APPLICATION NUMBER: US/09277, 716A  
EARLIER FILING DATE: 1999-03-26  
EARLIER APPLICATION NUMBER: 60/102, 939  
EARLIER FILING DATE: 1998-10-01  
EARLIER APPLICATION NUMBER: 60/089, 367  
EARLIER FILING DATE: 1998-06-15  
EARLIER APPLICATION NUMBER: 60/079, 624  
EARLIER FILING DATE: 1998-03-27  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 22  
LENGTH: 382  
TYPE: PRT  
ORGANISM: Artificial Sequence: fusion protein  
FEATURE:  
OTHER INFORMATION: Cellulose Binding Domain (CBD)-Gaussia luciferase fusion prote  
US-09-277-716-22  
Query Match 36.0%, Score 480.5; DB 4; Length 382;  
Best Local Similarity 62.6%, Pred. No. 4.1e-41;  
Matches 107; Conservative 10; Mismatches 33; Indels 21; Gaps 4;  
QY 1 MSVEFYNSNKSQNTSITPIIKITNTSDSLNLDNVKRYVYTSNCTGCFMCDHAGAL 60  
DB 1 MSVEFYNSNKSQNTSITPIIKITNTSDSLNLDNVKRYVYTSNCTGCFMCDHAGAL 60  
QY 61 LGNSVDMTSKYANFVKETASPTSTYDYLDPDSHMRCL--QGSSLIITISVFLVGAQ 117  
DB 61 LGNSVDMTSKYANFVKETASPTSTYDYLDPDSHMRCL--QGSSLIITISVFLVGAQ 117  
QY 118 NFSRO----EKGATGAVGVGAQQLGKSGGRVSMAGGA-VLGGLIGS 162  
DB 121 NTGTNDYSFPASSSTPVVNPRTGY-----IGCAKVLGTAPGS 159  
RESULT 2  
US-08-048-164A-2  
Sequence 2, Application US/08048164A  
Patent No. 5496934  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded

APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/048,164A  
APPLICATION NUMBER: US/08/048,164A  
FILING DATE: 14-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION/DOCKET NUMBER: 7809-003  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-048-164A-2

Query Match 35.7%; Score 477; DB 1; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSVEFYNSKNSAQTNSITPIIKITNTSDNLNDVKVRYRYYTSDGTGQTFWCDDHAGAL 60  
Db 6 MSVEFYNSKNSAQTNSITPIIKITNTSDNLNDVKVRYRYYTSDGTGQTFWCDDHAGAL 65  
Oy 61 LGNSYVDNTSKVTANFVKETASPTSTYDYL 92  
Db 66 LGNSYVDNTSKVTANFVKETASPTSTYDYL 97

RESULT 3  
US-08-460-462-2  
Sequence 2, Application US/08460462  
Patent No. 5670623  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,462  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/048,164  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION/DOCKET NUMBER: 7809-006  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-460-462-2

Query Match 35.7%; Score 477; DB 1; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSVEFYNSKNSAQTNSITPIIKITNTSDNLNDVKVRYRYYTSDGTGQTFWCDDHAGAL 60  
Db 6 MSVEFYNSKNSAQTNSITPIIKITNTSDNLNDVKVRYRYYTSDGTGQTFWCDDHAGAL 65  
Oy 61 LGNSYVDNTSKVTANFVKETASPTSTYDYL 92  
Db 66 LGNSYVDNTSKVTANFVKETASPTSTYDYL 97

RESULT 4  
US-08-460-457-2  
Sequence 2, Application US/08460457  
Patent No. 5719044  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
APPLICANT: Shpiegl, Itai  
APPLICANT: Goldstein, Marc A.  
APPLICANT: Dol, Roy H.  
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN FUSION PROTEINS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,457  
FILING DATE: concurrently herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/048,164  
FILING DATE: 14-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION/DOCKET NUMBER: 7809-008  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 162 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-460-457-2

Query Match 35.7%; Score 477; DB 1; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2,7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MSVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKVRYRYYTSDGTGOTFCMCDHAGAL 60  
Db 6 MSVEFYNSKSAQNTSTPIIKITNTSDSLNLDVKVRYRYYTSDGTGOTFCMCDHAGAL 65

Oy 61 LGNSYDNTSKVTANFVKETASPTSTYDTYLD 92  
Db 66 LGNSYDNTSKVTANFVKETASPTSTYDTYVE 97

RESULT 5  
US-08-460-458-2

; Sequence 2, Application US/08460458  
; Patent No. 5738984

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; TITLE OF INVENTION: KITS AND METHODS OF DETECTION USING CELLULOSE BINDING DOMAIN

; NUMBER OF SEQUENCES: 21

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,458

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

Oy 61 LGNSYDNTSKVTANFVKETASPTSTYDTYLD 92  
Db 66 LGNSYDNTSKVTANFVKETASPTSTYDTYVE 97

RESULT 6  
US-08-460-455-2

; Sequence 2, Application US/08460455

; Patent No. 5837814

; GENERAL INFORMATION:

; APPLICANT: Shoseyov, Oded

; APPLICANT: Shpiegl, Itai

; APPLICANT: Goldstein, Marc A.

; APPLICANT: Doi, Roy H.

; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN PROTEINS

; NUMBER OF SEQUENCES: 21

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,455

; FILING DATE: concurrently herewith

Tue Oct 29 08:31:12 2002

us-09-677-374-6.oligo.rai

APPLICANT: DOJ, Roy H.  
TITLE OF INVENTION: METHODS OF DETECTION USING THE  
CELLULOSE BINDING DOMAIN  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/330,394A  
FILING DATE: 27-OCT-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: CIP OF PCT/US94/04132  
FILING DATE: 14-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MISROCK, S. LESLIE  
REFERENCE/DOCKET NUMBER: 7809-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66441 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-330-394A-2

Query Match 35.7%; Score 477; DB 2; Length 162;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGQTFWCDHAGAL 60  
DB 6 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGQTFWCDHAGAL 65  
QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLD 92  
DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYVE 97

RESULT 8  
US-09-006-636-7  
Sequence 7, Application US/09006636  
Patent No. 6005092  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
ATTORNEY/AGENT INFORMATION:  
NAME: Shani, Ziv  
TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-  
GLUCANASE GENE, PROMOTER AND PROTEIN  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEO Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/006,636  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7809-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-006-636-7

Query Match 35.7%; Score 477; DB 3; Length 163;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGQTFWCDHAGAL 60  
DB 7 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKVRYRYYTSDGTGQTFWCDHAGAL 66  
QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLD 92  
DB 67 LGNSYVDNTSKVTANFVKETASPTSTYDYVE 98

RESULT 9  
US-09-006-632-7  
Sequence 7, Application US/09006632  
Patent No. 6184440  
GENERAL INFORMATION:  
APPLICANT: Shoseyov, Oded  
ATTORNEY/AGENT INFORMATION:  
NAME: Shigel, Etai  
TITLE OF INVENTION: TRANSGENIC PLANTS OF ALTERED  
MORPHOLOGY  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEO Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/006,632  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Baldwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7809-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-006-632-7

Query Match 35.7%; Score 477; DB 4; Length 163;  
Best Local Similarity 97.8%; Pred. No. 2.7e-41;  
Matches 90; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 60  
|||||  
DB 7 MSVEFYNSKSAQNTSITPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 66

QY 61 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92  
|||||  
DB 67 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 98

RESULT 10  
US-08-330-394A-29

Sequence 29, Application US/08330394A  
Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Yosef, Karmey

APPLICANT: Shpiegl, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: Dol, Roy H.

TITLE OF INVENTION: METHODS OF DETECTION USING THE

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/330,394A

FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CIP OF PCT/US94/04132

FILING DATE: 14-APR-1994

ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REFERENCE/DOCKET NUMBER: 7809-005

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66441 PENNIE

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 154 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-330-394A-29

Query Match 35.4%; Score 472; DB 2; Length 154;  
Best Local Similarity 97.8%; Pred. No. 8.2e-41;  
Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVEFYNSKSAQNTSITPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 61  
|||||  
DB 1 SVEFYNSKSAQNTSITPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 60

QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92  
|||||

DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 11

US-08-330-394A-22

Sequence 22, Application US/08330394A

Patent No. 5856201

GENERAL INFORMATION:

APPLICANT: Shoseyov, Oded

APPLICANT: Yosef, Karmey

APPLICANT: Shpiegl, Itai

APPLICANT: Goldstein, Marc A.

APPLICANT: Dol, Roy H.

TITLE OF INVENTION: METHODS OF DETECTION USING THE

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 AVENUE OF THE AMERICAS

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/330,394A

FILING DATE: 27-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CIP OF PCT/US94/04132

FILING DATE: 14-APR-1994

ATTORNEY/AGENT INFORMATION:

NAME: MISROCK, S. LESLIE

REFERENCE/DOCKET NUMBER: 7809-005

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66441 PENNIE

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-330-394A-22

Query Match 35.4%; Score 472; DB 2; Length 156;  
Best Local Similarity 97.8%; Pred. No. 8.4e-41;  
Matches 89; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVEFYNSKSAQNTSITPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 61  
|||||

DB 1 SVEFYNSKSAQNTSITPIIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 60

QY 62 GNSYVDNTSKVTANFVKETASPTSTYDTYLD 92  
|||||

DB 61 GNSYVDNTSKVTANFVKETASPTSTYDTYVE 91

RESULT 12

PCT-US95-13813-9

Sequence 9, Application PC/TUS9513813

GENERAL INFORMATION:

APPLICANT: Yeda Research and Development Co. Ltd.

APPLICANT: Ramot University Authority for Applied

APPLICANT: Research and Industrial Development Ltd.

APPLICANT: Bayer, Edward A.

APPLICANT: Moray, Ely

APPLICANT: Wilchek, Meli

```

; LENGTH: 493
; TYPE: PR
; ORGANISM: Bacillus licheniformis
; US-09-198-956-10

Query Match
Best Local Similarity 51.5%; Pred. No. 8.3e-17; Length 493;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSVEFYNSNKAQNTSTIPRIKITTSDSLNLNKKYKRYRYTSDGTGOTGFCMDHAGAL 60
   : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 335 LKVEFYNSNSDNTINSINPQFKYNTGSSAIDLKLTLLRYITVDGQXDGOTFMCDDHA-AI 393

QY 61 LGN--SYVDNISKVTANFVKETASTPSTDYLDPSHMRCLQ 101
   : | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 IGSNGSYNGITSNVKGTFFVMSSS-INNDYTLLEISFTGTL 435

RESULT 14
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: NO. 6187580el Pectate Lyases
; FILE REFERENCE: 5378-.200-US
; CURRENT APPLICATION NUMBER: US/09/198, 955A
; CURRENT FILING DATE: 1998-11-24
; PRIORITY FILING DATE: 1997-11-24
; PRIORITY APPLICATION NUMBER: 1344/97
; PRIORITY FILING DATE: 1997-11-24

```

```

? PRIOR FILING DATE: 1997-11-24
? PRIOR APPLICATION NUMBER: 60/067,249
? PRIOR FILING DATE: 1997-12-02
? PRIOR APPLICATION NUMBER: 60/067,240
? PRIOR FILING DATE: 1997-12-02
? PRIOR APPLICATION NUMBER: 09/073,684
? PRIOR FILING DATE: 1998-05-06
? PRIOR APPLICATION NUMBER: 09/184,217
? PRIOR FILING DATE: 1998-11-02
? NUMBER OF SEQ ID NOS: 32
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 12
? LENGTH: 493
? TYPE: prf
? ORGANISM: Clostridium thermocellum
US-09-198-935A-12

Query Match      18.4%   Score 245; DB 4; Length 493;
Best Local Similarity 51.5%; Pred.No. 8.3e-17;
Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3

QY    1 MSVEFYNSKSAQTNSTITPIIKITNTSDSLNLNRYKYRYTSDGQGOTFCWCDHAGAL 60
       :|:||||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    335 LKVEFYNSDPTJNSINPQFKYTNGSSADLSKLTLRIYYTYDGQKDQTFWCDDHA-AI 393
       :|:||||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY    61 LGN--SYVDNTSKYVTANFVETASPTSTDTYLDPFSHRGCLO 101
       :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    394 TGSNGSTNGCITSNVKGTFFVKMSSS-TNNADFTLEISFTGGTILE 435
       :|: || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-09-136-574A-43
Sequence 43, Application US/09136574A
Patent No. 6294366
```



```

1 GENERAL INFORMATION:
2 APPLICANT: Farrington, Graham K.
3 Anderson, Paige
4 Gibbs, Moreland
5 Bergquist, Peter
6 Daniels, Roy
7 Morgan, Hugh W.
8 Williams, Diane P.
9
10 TITLE OF INVENTION: Compositions and Methods for
11 Treating Cellulose Containing Fabrics Using Truncated
12 Cellulase Enzyme Compositions
13
14 NUMBER OF SEQUENCES: 49
15 CORRESPONDENCE ADDRESS:
16 ADDRESSEE: Howson and Howson
17 STREET: Spring House Corporate Center, P.O. Box 457
18 CITY: Spring House
19 STATE: PA
20 COUNTRY: USA
21 ZIP: 19477
22
23 COMPUTER READABLE FORM:
24 MEDIUM TYPE: Diskette
25 COMPUTER: IBM Compatible
26 OPERATING SYSTEM: DOS
27
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/09/136,574A
30 FILING DATE: 19-Aug-1998
31
32 CLASSIFICATION: <Unknown>
33
34 PRIOR APPLICATION DATA:
35 APPLICATION NUMBER: US 08/932,571
36 FILING DATE: September 19, 1997
37
38 ATTORNEY/AGENT INFORMATION:
39 NAME: Bak, Mary E.
40 REGISTRATION NUMBER: 31,215
41 REFERENCE/DOCKET NUMBER: 1997US001/CIP
42
43 TELECOMMUNICATION INFORMATION:
44 TELEPHONE: 215-540-9200
45 TELEFAX: 215-540-5818
46
47 TELEX: <Unknown>
48
49 INFORMATION FOR SEQ ID NO: 43:
50 SEQUENCE CHARACTERISTICS:
51 LENGTH: 1426 amino acids
52 TYPE: amino acid
53 STRANDEDNESS: single
54 TOPOLOGY: linear
55 MOLECULE TYPE: No. 6294366e
56 SEQUENCE DESCRIPTION: SEQ ID NO: 43:
57
58 US-09-136-574A-43
59
60 Query Match 11.2%; Score 149.5; DB 4; Length 1426;
61 Best Local Similarity 32.9%; Pred. No. 2.4e-06;
62 Matches 50; Conservative 16; Mismatches 35; Indels 51; Gaps 6;
63
64 Oy 1 MSVEFYNSNKSNOTNSIPRIKITNTSSDLNLDVKKRYRYSDDGTGGTFWCDHA--G 58
65 Db 414 LKVLTKNNETASIGSIRPMKPIVINGSSSYDLSKVKIRIWTYVDGDKPQSNVCDMAIG 473
66 Oy 59 ALLGNSYVDNTSKYTANFVKETASPTSTYYDLYLDPSHRMGLQGSLLIISVFLGCAQN 118
67 Db 474 A-----SNVTFNFVK--LSSGSGADAYL----- 495
68 Oy 119 FSRQEVGAATGAVVGVAGOLF-GKSGSRVSN 149
69 Db 496 ---EVGFSSG-----AGQLPGKADADIDV 517

```

Search completed: October 27, 2002, 11:12:14  
Job time : 14.2642 secs

**THIS PAGE BLANK (USPTO)**



PA (BURI/) BURIAN J.  
PA (KUZY/) KUZYK M A.  
XX  
XX  
PI Kay WW, Burian J, Kuzyk MA;  
PI  
XX WPI: 2001-316844/34.  
XX  
DR N-PSDI: AAF86248.  
DR  
XX  
XX Method for protecting poikilothermic fish against salmonid rickettsial  
XX  
XX septicaemia and other rickettsial diseases comprises administering a  
XX  
XX vaccine containing the OSPA protein of *Piscirickettsia salmonis* -  
XX  
XX  
XX Example 4; Fig 5; 35pp: English

This invention relates to a method for the protection against infection of a polkilothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of a *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of OspA in the form of a vaccine. The method is used for protecting animals particularly polkilothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmonid rickettsial septicemia (SRS) and other rickettsial diseases. The present sequence represents the amino acid sequence of C17E2, a *P. salmonis* OspA construct optimised for expression in *Escherichia coli*, fused to an undefined N-terminal fusion partner. The fusion protein is used in a vaccine to create an anti-OspA antibody response.

Query Match	Best Local Similarity	100.0%;	Score 1335;	DB 22;	Length 256;
Matches	256;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	
Qy	1	MSVEFYNSKSKAQTNSITPIIKITINTSDPLNDYKRYVYTYTSDGTGGCFWCDHAGAL	60		
Db	1	MSVEFYNSKSKAQTNSITPIIKITINTSDPLNDYKRYVYTYTSDGTGGCFWCDHAGAL	60		
Qy	61	LGNSYVUNYSKYATNVKETAAPTSTYDYLDPSHMRGCLQGSSLLITSVFLVCAQNF	120		
Db	61	LGNSYVUNYSKYATNVKETAAPTSTYDYLDPSHMRGCLQGSSLLITSVFLVCAQNF	120		
Qy	121	ROEVAATGAVGAVAGAGOLFEGKSGRSVMAIGCAVLGLIGSKRIGQSMDOODKIKLNSL	180		
Db	121	ROEVAATGAVGAVAGAGOLFEGKSGRSVMAIGCAVLGLIGSKRIGQSMDOODKIKLNSL	180		
Qy	181	EKKVAGQYTRMRNDPTGNSYSVEPVRTYQRPKNKERROQYCRFPOOKAMIAQOKETIYGT	240		
Db	181	EKKVAGQYTRMRNDPTGNSYSVEPVRTYQRPKNKERROQYCRFPOOKAMIAQOKETIYGT	240		
Qy	241	ACFPDGRKQVISTEK	256		
Db	241	ACFPDGRKQVISTEK	256		

RESULT 2  
AAB81127  
TD AAB81127 standard; Protein: 161 AA

AA	11-JUL-2001 (first entry)	17E2 amino acid sequence
DT		
XX		

XX piscirickettsia salmonis; rickettsial pathogen, poikilothermic fish; rickettsial disease  
KW rickettsial septicemia; rickettsial disease

KW vaccine; ospa; salmonella  
KW SRS.

XX piscirickettsia salmonis.  
OS

Synthetic.

XX	XX	Location/Qualifiers
XX	XX	XX

Key	Region
FH	
FT	

3.

3

FT		1980-1981
XX	CA28281913-A1.	
PN		
XX	17-MAR-2001.	
PD		
XX	17-SEP-1999;	99CA-2281913.
PF		
XX	17-SEP-1999;	99CA-2281913
PR		

XX	(KAYW/)	KAY W. W.
PA	(BURI)	BURIAN J.
PA	(KUZY/)	KUZYK M. A.
XX		
PI	Kay WW,	Burian J, Kuzyk MA;
XX		
WPI	2001-316844/34.	
DR	N-PSDB; AAF86247.	
DR		

XX Method for protecting poikilothermic fish against salmonid rickettsial  
PT septicemia and other rickettsial diseases comprises administering a  
PT vaccine containing the OSPA protein of *Piscirickettsia salmonis* -  
XX  
PS Disclosure: Fig 5; 35pp: English.

XX This invention relates to a method for the protection against infection of a polkitlothermic fish by the bacterial pathogen, *Piscirickettsia salmonis*. The method comprises administering an immunogenic amount of a *P. salmonis* specific antigen termed OspA, or an immunogenic fragment of OspA in the form of a vaccine. The method is used for protecting animals, particularly polkitlothermic fish, against the bacterial pathogen *P. salmonis*. The method is also useful for protecting against salmonid rickettsial septicaemia (SRS) and other rickettsial diseases. The present invention represents optimised *P. salmonis* OspA protein 17E2. The DNA sequence encoding optimised *P. salmonis* OspA protein 17E2, the DNA encoding OspA 17E2 (AAFP6247) has been optimised for expression in *Escherichia coli*. An OspA protein with an N-terminal fusion partner is used in a vaccine to create an anti-OspA antibody response.

xx	Sequence	161 AA;	62.3%;	Score 832;	DB 22;	Length 161;
sq	Query Match					
	Best Local Similarity	100.0%;	Pred. No. 1 6e-73;			
	Matches 161;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0
QY	96	MRCGLGSSLLIIITSEFLVCAQNFESRQEGVATGAVGVGAVGOLFGKGGSGRVSMATIGAV	155			
Db	1	MRCGLGSSLLIIITSEFLVCAQNFESRQEGVATGAVGVGAVGOLFGKGGSGRVSMATIGAV	60			
QY	156	LGGLISKTIQSDHDDDKTKLNLSLEKAVAGOVTRMRPDPDTSNSYSEVPRTYORKOE	215			
Db	61	LGGLISKTIQSDHDDDKTKLNLSLEKAVAGOVTRMRPDPDTSNSYSEVPRTYORKOE	120			
QY	216	RPODYCREFOQKAMIAQKOEIYGTACPOPDRMVOYSTEK	256			
Db	121	RPODYCREFOQKAMIAQKOEIYGTACPOPDRMVOYSTEK	161			

RESULT 3  
AAG78025  
ID AAG78025 standard; protein; 162 AA

AC AAG78025;  
XX  
DT 15-JAN-2002 (first entry)

biochem. J. 1987, 241, 115-120.

DE piscirickettsiosis; salmonid rickettsial  
XX disease; salmonid fish: piscirickettsiosis; salmonid rickettsial

KW Placitricellusidae; surface antigen; vaccine; antibacterials; septicemia; SRS

KW ATCC VR-1361-

XX  
OS Piscirickettsia salmonis.

XX

PN WO200168865-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 12-MAR-2001; 2001WO-GB01055.  
 XX  
 PR 11-MAR-2000; 2000GB-0005838.  
 PR 01-JUL-2000; 2000GB-0016080.  
 PR 01-JUL-2000; 2000GB-0016082.  
 PR 29-JUL-2000; 2000GB-0018599.  
 XX  
 PA (AQUA-) AQUA HEALTH EURO LTD.  
 XX  
 PI Sismard N. Brouwers H, Jones S, Griffiths S, Valenzuela P;  
 PI Burzio L;  
 DR WPI: 2001-639050/73.  
 DR N-PSDB; AAH79040.  
 XX  
 PT New nucleic acids encoding an amino acid sequence homologous to the  
 PT surface antigen present on Piscirickettsia salmonis are useful to  
 PT protect fish against Piscirickettsiosis -  
 XX  
 PS Claim 6; Fig 5; 25pp; English.  
 XX  
 CC The invention relates to nucleic acid sequences and the encoded protein  
 CC of a least part of the surface antigen present on Piscirickettsia  
 CC salmonis for production of a vaccine with antibacterial activity to  
 CC protect fish against P. salmonis which causes piscirickettsiosis, also  
 CC known as salmonid rickettsial septicaemia.  
 CC  
 SQ Sequence 162 AA:

Query Match 61.0%; Score 815; DB 22; Length 162;  
 Best Local Similarity 98.8%; Pred. No. 7.6e-72;  
 Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 97 RGCLOGSSLIITISVFLVGCACNFSROEYGAATGAVGVAGOLFSGKSGRYMAIGAVL 156  
 |||||||  
 Db 3 RGCLOGSSLIITISVFLVGCACNFSROEYGAATGAVGVAGOLFSGKSGRYMAIGAVL 62  
 QY 157 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGSNSYSVEPVRTYQRYNKOER 216  
 |||||||  
 Db 63 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGSNSYSVEPVRTYQRYNKOER 122  
 QY 217 RQYCYREFQOKAMITAGOKOEIYGTACPODGRMOWISTEK 256  
 |||||||  
 Db 123 RQYCYREFQOKAMITAGOKOEIYGTACPODGRMOWISTEK 162

RESULT 4  
 AAB81126  
 ID AAB81126 standard; Protein; 162 AA.  
 XX  
 AC AAB81126;  
 XX  
 DT 11-JUL-2001 (first entry)  
 XX  
 DE Ospa antigen amino acid sequence.  
 XX  
 KM Polkilothemic fish; Piscirickettsia salmonis; rickettsial pathogen;  
 KM vaccine; Ospa; salmonid rickettsial septicaemia; rickettsial disease;  
 KM SRS.  
 XX  
 OS Piscirickettsia salmonis.  
 XX  
 FH Key Location/Qualifiers  
 FT 110..129  
 FT Region /label= B\_cell-epitope  
 PN CA2281913-A1.  
 XX  
 PD 17-MAR-2001.

XX  
 PF 17-SEP-1999; 99CA-2281913.  
 XX  
 PR 17-SEP-1999; 99CA-2281913.  
 XX  
 PA (KAYW/) KAY W W.  
 PA (BURI/) BURIAN J.  
 PA (KUZV/) KUZV M A.  
 XX  
 PI Kay WM, Burian J, Kuzv MA;  
 XX  
 DR WPI: 2001-316844/34.  
 DR N-PSDB; AAF86246.  
 XX  
 PT Method for protecting polkilothemic fish against salmonid rickettsial  
 PT septicaemia and other rickettsial diseases comprises administering a  
 PT vaccine containing the Ospa protein of Piscirickettsia salmonis -  
 XX  
 PS Example 2; Fig 2B; 35pp; English.  
 XX  
 CC This invention relates to a method for the protection against infection  
 CC of a polkilothemic fish by the bacterial pathogen, Piscirickettsia  
 CC salmonis. The method comprises administering an immunogenic fragment of a  
 CC P. salmonis specific antigen termed Ospa, or an immunogenic fragment of  
 CC Ospa in the form of a vaccine. The method is used for protecting animals,  
 CC particularly polkilothemic fish, against the bacterial pathogen  
 CC P. salmonis. The method is also useful for protecting against salmonid  
 CC rickettsial septicaemia (SRS) and other rickettsial diseases. The present  
 CC sequence represents P. salmonis Ospa protein. An Ospa protein with an  
 CC N-terminal fusion partner is used in a vaccine to create an anti-Ospa  
 CC antibody response.  
 CC  
 SQ Sequence 162 AA:

Query Match 61.0%; Score 815; DB 22; Length 162;  
 Best Local Similarity 98.8%; Pred. No. 7.6e-72;  
 Matches 158; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 97 RGCLOGSSLIITISVFLVGCACNFSROEYGAATGAVGVAGOLFSGKSGRYMAIGAVL 156  
 |||||||  
 Db 3 RGCLOGSSLIITISVFLVGCACNFSROEYGAATGAVGVAGOLFSGKSGRYMAIGAVL 62  
 QY 157 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGSNSYSVEPVRTYQRYNKOER 216  
 |||||||  
 Db 63 GGLIGSKIGOSMDQDDIKLNQSLKVKAGOVTRMRNDTGSNSYSVEPVRTYQRYNKOER 122  
 QY 217 RQYCYREFQOKAMITAGOKOEIYGTACPODGRMOWISTEK 256  
 |||||||  
 Db 123 RQYCYREFQOKAMITAGOKOEIYGTACPODGRMOWISTEK 162

RESULT 5  
 AAE05746  
 ID AAE05746 standard; Protein; 190 AA.  
 XX  
 AC AAE05746;  
 XX  
 DT 24-SEP-2001 (first entry)  
 XX  
 DE Clostridium cellulovorans cellulose binding domain-180 (CBD-180).  
 XX  
 KM Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
 KM yarn; fiber; textile; biological crosslinker; mechanical property;  
 KM Clostridium cellulovorans cellulose binding domain; wet strength;  
 KM durability; elasticity; CBDc10s; cellulose binding protein A; CBP A;  
 KM CBD-180.  
 XX  
 OS Clostridium cellulovorans.  
 XX  
 PN WO200134091-A2.  
 XX  
 PD 17-MAY-2001.

```

PF 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
PR 18-NOV-1999; 99US-0166389.
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX Levy I, Nussinovitch A, Shoseyov O;
XX WPI: 2001-457121/49.
DR N-PSDB; AAD11043.
XX Preparation of a polysaccharide containing material having at least one
PT desired structural, chemical, physical, electrical and/or mechanical
PT property -
XX Example 1.2; Fig 1e-1g; 121pp; English.
XX The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
CC The method involves treating the polysaccharide structure with a
CC polysaccharide binding domain (PBD) fusion protein. The method is
CC used to alter the structural, chemical, physical, electrical and
CC mechanical properties of polysaccharide materials such as paper,
CC yarns, fibers and textiles; using biological crosslinking agents.
CC The polysaccharide containing materials have improved mechanical
CC properties such as wet strengths, durability and elasticity. The PBD
CC reagent is applied in the forming stage in fluting paper manufacture
CC which eliminates the sizing step. The use of a biological crosslinker
CC improves the recyclability of paper products. The PBD reagent maintains
CC the fine fibers in a slurry therefore resulting in better recovery of
CC raw materials. The PBD molecules are eluted by strong alkaline conditions
CC which enhances the ability of the alkaline glue used in binding to
CC penetrate paper.
CC The present sequence is Clostridium cellulovorans cellulose binding
CC domain (CBD)-180. CBD-180 contains 180 amino acids of C. cellulovorans
CC CBD (CBDClos) of cellulose binding protein A (CBP A).
XX Sequence 190 AA:
SQ Query Match 36.5%; Score 487; DB 22; Length 190;
Best Local Similarity 46.8%; Pred. No. 1.4e-39;
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;
QY 1 MSVEFYNSKSAQNSITPIIKITNTSDSLNDLVKVRYYTSDGTGQTFWCHAGAL 60
DB 1 MSVEFYNSKSAQNSITPIIKITNTSDSLNDLVKVRYYTSDGTGQTFWCHAGAL 60
QY 61 LGNSYVDNTSVYTNFVKETASPTSTYDTYDPSHMGCLQGSLLITISVFLGCAQNF 120
DB 61 LGNSYVDNTSVYTNFVKETASPTSTYDTYDPSHMGCLQGSLLITISVFLGCAQNF 120
QY 121 RQEVGAATGAVGAVGAGQLFGKSGRVSMAIGAVLGLIGSKIGSGMDQDKIKLNQSL 180
DB 121 RQEVGAATGAVGAVGAGQLFGKSGRVSMAIGAVLGLIGSKIGSGMDQDKIKLNQSL 180
QY 93 -----FGPASGRATL-----KKGQFITTQ----- 111
DB 93 -----FGPASGRATL-----KKGQFITTQ----- 111
QY 181 EKVKAGQVTR-WRNPDTGNSTSYVEPVTRYQRYNKKQHRQOYCREFQOKAMIAQOKO 238
DB 181 EKVKAGQVTR-WRNPDTGNSTSYVEPVTRYQRYNKKQHRQOYCREFQOKAMIAQOKO 238
QY 112 -----GRITKSDMSMYTOTNDISFDASSSTPVNP-----KYTGIGGAK--VL 153
DB 112 -----GRITKSDMSMYTOTNDISFDASSSTPVNP-----KYTGIGGAK--VL 153
QY 239 GRACPOPD 246
DB 239 GRACPOPD 246
QY 154 GTA-PGPD 160
DB 154 GTA-PGPD 160
RESULT 6
ID AAE05749 standard: Protein; 328 AA.
XX AAE05749;
XX AAE05749;
XX 24-SEP-2001 (first entry)

```

```

XX Chimeric S peptide-cellulose binding domain-S protein.
XX Polysaccharide modification; polysaccharide binding domain; PBD; paper;
XX yarn; fiber; textile; biological crosslinker; mechanical property;
XX wet strength; durability; elasticity; cellulose binding domain; CBD;
XX chimeric protein; S peptide-cellulose binding domain-S protein;
XX Sped-CBD-Sprot; bovine.
XX Chimeric - Clostridium cellulovorans.
XX Chimeric - Bos sp.
XX Key location/qualifiers
XX Region 30..208
XX Region 226..326
XX Region 327
XX Misc-difference 327
XX /label="Unknown
XX /note="Encoded by TAG"
XX Misc-difference 327..328
XX /note="These residues are absent in the sequence shown
XX as SEQ ID NO: 10 in the sequence listing"
XX WO200134091-A2.
XX 17-MAY-2001.
XX 02-NOV-2000; 2000MO-IL00708.
XX 08-NOV-1999; 99US-0164140.
XX 18-NOV-1999; 99US-0166389.
XX (CBDT-) CBD TECHNOLOGIES LTD.
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX Levy I, Nussinovitch A, Shoseyov O;
XX WPI: 2001-457121/49.
XX N-PSDB; AAD11046.
XX Preparation of a polysaccharide containing material having at least one
XX desired structural, chemical, physical, electrical and/or mechanical
XX property -
XX Example 4; Fig 4b-4g; 121pp; English.
XX The present invention relates to methods and compositions for cross-
XX linking and/or modifying the properties of polysaccharide materials.
XX The method involves treating the polysaccharide structure with a
XX polysaccharide binding domain (PBD) fusion protein. The method is
XX used to alter the structural, chemical, physical, electrical and
XX mechanical properties of polysaccharide materials such as paper,
XX yarns, fibers and textiles; using biological crosslinking agents.
XX The polysaccharide containing materials have improved mechanical
XX properties such as wet strengths, durability and elasticity. The PBD
XX reagent is applied in the forming stage in fluting paper manufacture
XX which eliminates the sizing step. The use of a biological crosslinker
XX improves the recyclability of paper products. The PBD reagent maintains
XX the fine fibers in a slurry therefore resulting in better recovery of
XX raw materials. The PBD molecules are eluted by strong alkaline conditions
XX which enhances the ability of the alkaline glue used in binding to
XX penetrate paper.
XX The present sequence is S peptide-cellulose binding domain-S protein
XX (Sped-CBD-Sprot), a fusion protein derived from Clostridium cellulovorans
XX and bovine.
XX Note: This sequence is stated to be the same as that shown as
XX SEQ ID NO: 10 in the sequence listing of the specification. However
XX this sequence has 2 additional residues at its C-terminal.
XX Sequence 328 AA:
SQ Query Match 36.5%; Score 487; DB 22; Length 328;

```

Best Local Similarity 46.8%; Pred. No. 2.9e-39;  
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

QY 1 MSVEFYNSKNSAQTNSTITPIIKITNTSDSLNDLVKVRYYRYSYDGTGOTFWCDHAGAL 60  
DB 35 MSVEFYNSKNSAQTNSTITPIIKITNTSDSLNDLVKVRYYRYSYDGTGOTFWCDHAGAL 94  
QY 61 LGSNSYDNTSKVTANFKEKETASPTSTYDTLDPSSHMGCLGGSSLIITSFVLVCAQNFNS 120  
DB 95 LGSNSYDNTSKVTANFKEKETASPTSTYDTLDPSSHMGCLGGSSLIITSFVLVCAQNFNS 126  
QY 121 ROEVGAATGAVVGAVAGLFGKSGRYVSMAGAVLGLGSKIGQSMDOODKIKLNQSL 180  
DB 127 -----GRTKSDMSNYQTNDYSFDASSSTPPVNP-----KKGQRTTIQ----- 145  
QY 181 EKVKAGQVTR--WRNPDTGNSYSVEPVRYORYNKORROOYCREFOQKAMIAQOKEIY 238  
DB 146 -----GRTKSDMSNYQTNDYSFDASSSTPPVNP-----KVTGYIGGAK--VL 187  
QY 239 GTACPOPD 246  
DB 188 GTA-PCPD 194

RESULT 7  
ID AAE05747 standard; Protein; 341 AA.  
AC AAE05747;

24-SEP-2001 (first entry)

DE Clostridium cellulovorans CBD cross linker protein (CCP).

KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
KM yarn; fiber; textile; biological crosslinker; mechanical property;  
KW wet strength; durability; elasticity; cellulose binding domain; CBD;  
KM CBD cross linker protein; CCP.

OS Clostridium cellulovorans.

FX Key Location/Qualifiers

FT Misc-difference 340..341

FT /note- "Encoded by CCATAGGAT"

FT Misc-difference 341  
FT /note- "This residue is absent in the sequence shown  
as SEQ ID NO: 6 in the sequence listing"

PN MO200134091-A2.

PD 17-MAY-2001.

PF 02-NOV-2000; 2000MO-IL00708.

PR 08-NOV-1999; 99US-0164140.

PR 18-NOV-1999; 99US-0166389.

PA (CBDT-) CBD TECHNOLOGIES LTD.

PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

PI Levy I, Nussimovitch A, Shoseyov O;

DR WPI; 2001-457121/49.

DR N-PSDB; AAD11044.

PT Preparation of a polysaccharide containing material having at least one  
desired structural, chemical, physical, electrical and/or mechanical  
property

Example 1.2; Fig 2b-2e; 121pp; English.

CC The present invention relates to methods and compositions for cross-  
linking and/or modifying the properties of polysaccharide materials.

CC The method involves treating the polysaccharide structure with a  
CC polysaccharide binding domain (PBD) fusion protein. The method is  
CC used to alter the structural, chemical, physical, electrical and  
CC mechanical properties of polysaccharide materials such as paper,  
CC yarns, fibers and textiles, using biological crosslinking agents.  
CC The polysaccharide containing materials have improved mechanical  
CC properties such as wet strength, durability and elasticity. The PBD  
CC reagent is applied in the forming stage in fluting paper manufacture  
CC which eliminates the sizing step. The use of a biological crosslinker  
CC improves the recyclability of paper products. The PBD reagent maintains  
CC the fine fibers in a slurry therefore resulting in better recovery of  
CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
CC which enhances the ability of the alkaline glue used in binding to  
CC penetrate paper.

CC The present sequence is Clostridium cellulovorans cellulose binding

CC domain (CBD) cross linker protein (CCP).

CC Note: This sequence is stated to be the same as that shown as

CC SEQ ID NO: 6 in the sequence listing of the specification. However

CC this sequence has an additional residue at its C-terminal.

QY 1 MSVEFYNSKNSAQTNSTITPIIKITNTSDSLNDLVKVRYYRYSYDGTGOTFWCDHAGAL 60  
DB 1 MSVEFYNSKNSAQTNSTITPIIKITNTSDSLNDLVKVRYYRYSYDGTGOTFWCDHAGAL 60

QY 61 LGSNSYDNTSKVTANFKEKETASPTSTYDTLDPSSHMGCLGGSSLIITSFVLVCAQNFNS 120  
DB 61 LGSNSYDNTSKVTANFKEKETASPTSTYDTLDPSSHMGCLGGSSLIITSFVLVCAQNFNS 120

QY 121 ROEVGAATGAVVGAVAGLFGKSGRYVSMAGAVLGLGSKIGQSMDOODKIKLNQSL 180  
DB 121 ROEVGAATGAVVGAVAGLFGKSGRYVSMAGAVLGLGSKIGQSMDOODKIKLNQSL 180

QY 181 EKVKAGQVTR--WRNPDTGNSYSVEPVRYORYNKORROOYCREFOQKAMIAQOKEIY 238  
DB 181 EKVKAGQVTR--WRNPDTGNSYSVEPVRYORYNKORROOYCREFOQKAMIAQOKEIY 238

QY 239 GTACPOPD 246  
DB 239 GTACPOPD 246

QY 154 GTA-PCPD 160  
DB 154 GTA-PCPD 160

RESULT 8  
ID AAR63634 standard; Protein; 162 AA.  
AC AAR63634;

05-JUN-1995 (first entry)

DE Cellulose binding domain.

FX Cellulose binding domain; CBD.

OS Clostridium cellulovorans.

PN MO9424158-A.

PD 27-OCT-1994.

PF 14-APR-1994; 94WO-US04132.

PR 14-APR-1993; 93US-0048164.

PA (REGC) UNIV CALIFORNIA.

PA (YISS) YISSUM RES & DEV CO.

PI Dol RH, Goldstein MA, Shoseyov O, Shplegl I;

XX DR WPI; 1994-341767/42.  
 XX N-PSDB; AAO72917.  
 XX PT Isolated cellulose binding domain and fusion proteins - with  
 PT applications, including drug delivery, affinity separations, and  
 PT diagnostic techniques  
 XX PS Claim 1; Fig 1; 125pp; English.  
 XX CC AAO72917/R63634 is a novel isolated cellulose binding domain. It pref.  
 CC binds crystalline with a Kd ranging from 1.5 microm - 0.5 microm,  
 CC pref. with a Kd of <1.2 microm, esp. <1.0 microm. The protein and  
 CC the encoding DNA are claimed, as is a CBD fusion protein (FP)  
 CC comprising the CBD and a second protein. The second protein is pref.  
 CC Protein A, heparinase, or a hormone or an enzyme capable of degrading  
 CC an environmental pollutant, or an HSP, HSP antibody, cross-reactive  
 CC HSP-related protein or an antigenic portion of this. The CBD and FP  
 CC may be used in drug delivery, affinity seps, and diagnostic  
 CC techniques. CBD nucleic acid may be obtd. from a variety of cell  
 CC sources that produce CBDs that bind with high affinity and in a  
 CC reversible manner or that produce CBD encoding mRNA. The preferred  
 CC source of CBD encoding nucleic acid is C. cellulovorans.  
 XX CC  
 XX Sequence 162 AA:  
 XX  
 XX Query Match 36.2%; Score 483; DB 15; Length 162;  
 XX Best Local Similarity 51.7%; Pred. No. 2.7e-39;  
 XX Matches 106; Conservative 9; Mismatches 14; Indels 76; Gaps 4;  
 QY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYYTSGTGQGFMCDHAGAL 60  
 DB 6 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYYTSGTGQGFMCDHAGAL 65  
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPESHMRGCLQGSSLIISVFLVGCANFMS 120  
 DB 66 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPESHMRGCLQGSSLIISVFLVGCANFMS 120  
 QY 121 ROEYGAATGAVGVAGQLFGKSGRYSMAIGAVLGLIGSKIGOSMDQDKIKLNSL 180  
 DB 98 -----FGFASGRATL-----KKGQFPIIQ----- 116  
 QY 181 EKVRAGQVTR--WRNPDTGNSYSVE 203  
 DB 117 -----GRITKSDMSNYQTNDYSFD 136  
 RESULT 9  
 ID AAE05745  
 AC AAE05745 standard; Protein: 163 AA.  
 XX AC AAE05745;  
 XX DT 24-SEP-2001 (first entry)  
 XX DE Clostridium cellulovorans cellulose binding domain (CBDclos).  
 XX KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
 KW yarn; fiber; textile; biological crosslinker; mechanical property;  
 KW Clostridium cellulovorans cellulose binding domain; wet strength;  
 KW durability; elasticity; CBDclos; cellulose binding protein A; CBP A.  
 XX OS Clostridium cellulovorans.  
 XX PN WO200134091-A2.  
 XX PD 17-MAY-2001.  
 XX PF 02-NOV-2000; 2000WO-IL00708.  
 XX PR 08-NOV-1999; 99US-0164140.  
 XX PR 18-NOV-1999; 99US-0166389.  
 XX XX

PA (CBDT-) CBD TECHNOLOGIES LTD.  
 PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
 XX PI Levy I, Nussinovitch A, Shoseyov O;  
 XX DR WPI; 2001-457121/49.  
 XX N-PSDB; AAD11042.  
 XX PT Preparation of a polysaccharide containing material having at least one  
 PT desired structural, chemical, physical, electrical and/or mechanical  
 PT property  
 XX PS Example 1.1; Page 111-112; 121pp; English.  
 XX CC The present invention relates to methods and compositions for cross-  
 CC linking and/or modifying the properties of polysaccharide materials.  
 CC The method involves treating the polysaccharide structure with a  
 CC polysaccharide binding domain (PBD) fusion protein. The method is  
 CC used to alter the structural, chemical, physical, electrical and  
 CC mechanical properties of polysaccharide materials such as paper,  
 CC yarns, fibers and textiles, using biological crosslinking agents.  
 CC The polysaccharide containing materials have improved mechanical  
 CC properties such as wet strengths, durability and elasticity. The PBD  
 CC reagent is applied in the forming stage in fluting paper manufacture  
 CC which eliminates the sizing step. The use of a biological crosslinker  
 CC improves the recyclability of paper products. The PBD reagent maintains  
 CC the fine fibers in a slurry therefore resulting in better recovery of  
 CC raw materials. The PBD molecules are eluted by strong alkaline conditions  
 CC which enhances the ability of the alkaline glue used in binding to  
 CC penetrate paper.  
 CC The present sequence is Clostridium cellulovorans cellulose binding  
 CC domain (CBDclos) of cellulose binding protein A (CBP A).  
 XX CC  
 XX Sequence 163 AA:  
 XX  
 XX Query Match 36.2%; Score 483; DB 22; Length 163;  
 XX Best Local Similarity 51.7%; Pred. No. 2.7e-39;  
 XX Matches 106; Conservative 9; Mismatches 14; Indels 76; Gaps 4;  
 QY 1 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYYTSGTGQGFMCDHAGAL 60  
 DB 7 MSVEFYNSKSAQTNSITPIIKITNTSDSLNDLVKRYRYYTSGTGQGFMCDHAGAL 66  
 QY 61 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPESHMRGCLQGSSLIISVFLVGCANFMS 120  
 DB 67 LGNSYVDNTSKVTANFVKETASPTSTYDYLDPESHMRGCLQGSSLIISVFLVGCANFMS 120  
 QY 121 ROEYGAATGAVGVAGQLFGKSGRYSMAIGAVLGLIGSKIGOSMDQDKIKLNSL 180  
 DB 99 -----FGFASGRATL-----KKGQFPIIQ----- 117  
 QY 181 EKVRAGQVTR--WRNPDTGNSYSVE 203  
 DB 118 -----GRITKSDMSNYQTNDYSFD 137  
 RESULT 10  
 ID AAE05748  
 AC AAE05748 standard; Protein: 428 AA.  
 XX AC AAE05748;  
 XX DT 24-SEP-2001 (first entry)  
 XX DE Clostridium cellulovorans protein A-cellulose binding domain.  
 XX KW Polysaccharide modification; polysaccharide binding domain; PBD; paper;  
 KW yarn; fiber; textile; biological crosslinker; mechanical property;  
 KW wet strength; durability; elasticity; protein A-cellulose binding domain;  
 KW ProtA-CBD.  
 XX OS Clostridium cellulovorans.  
 XX XX



FH	Key	Location/Qualifiers
FT	Region	1..263
FT		/label= Protein_A
FT		/note= "This region is derived from PRIT2T vector"
FT	Region	265..426
FT		/label= CBP_A-protein
FT	Misc-difference	427
FT		/label= Unknown
FT		/note= "Encoded by TAG"
FT	Misc-difference	427..428
FT		/note= "These residues are absent in the sequence shown as SEQ ID NO: 8 in the sequence listing"
XX		
PX	MO200134091-A2.	
XX		
PD	17-MAY-2001.	
XX		
PF	02-NOV-2000; 2000WO-IL00708.	
XX		
PR	08-NOV-1999; 99US-0164140.	
PR	18-NOV-1999; 99US-0166389.	
XX		
PA	(CBDT-) CBD TECHNOLOGIES LTD.	
PA	(YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.	
XX		
PI	Levy I, Nussinovitch A, Shoseyov O;	
XX		
DR	WPI: 2001-457121/49.	
DR	N-PSDB: MAD11045.	
XX		
PT	Preparation of a polysaccharide containing material having at least one desired structural, chemical, physical, electrical and/or mechanical property -	
XX		
PS	Example 1.3; Fig 3b-3g; 121pp; English.	
XX		
CC	The present invention relates to methods and compositions for cross-	
CC	linking and/or modifying the properties of polysaccharide materials.	
CC	The method involves treating the polysaccharide structure with a	
CC	polysaccharide binding domain (PBD) fusion protein. The method is	
CC	used to alter the structural, chemical, physical, electrical and	
CC	mechanical properties of polysaccharide materials such as paper,	
CC	yarns, fibers and textiles, using biological crosslinking agents.	
CC	The polysaccharide containing materials have improved mechanical	
CC	properties such as wet strengths, durability and elasticity. The PBD	
CC	reagent is applied in the forming stage in fluidizing paper manufacture	
CC	which eliminates the sizing step. The use of a biological crosslinker	
CC	improves the recyclability of paper products. The PBD reagent maintains	
CC	the fine fibers in a slurry therefore resulting in better recovery of	
CC	raw materials. The PBD molecules are eluted by strong alkaline conditions	
CC	which enhances the ability of the alkaline glue used in binding to	
CC	penetrate paper.	
CC	The present sequence is Clostridium cellulovorans Protein A-cellulose	
CC	binding domain (ProCA-CBD). This sequence contains a part of CBP A	
CC	protein and a part of protein A derived from pRIT2T vector sequence.	
CC	Note: In the sequence listing of the specification. However	
CC	SEQ ID NO: 8 in the sequence listing of the specification. However	
CC	this sequence has 2 additional residues at its C-terminal.	
XX		
SQ	Sequence 428 AA:	
	Query Match 36.2%; Score 483; DB 22; Length 428;	
	Best Local Similarity 51.7%; Pred. No. 1e-38; Mismatches 14; Indels 76; Gaps	
	Matches 106; Conservative 9;	
OY	1 MSVEFYNSNKSQTNSTPIIKITRTSDSLNLNDVKRYYYVTSOGTGQTFWCDAHAL 60	
Db	270 MSVEFYNSNKSQTNSTPIIKITRTSDSLNLNDVKRYYYVTSOGTGQTFWCDAHAL 329	
OY	61 LGNSVDVMTSKVTAFNVETASPTSTPYDPLDPHSIMRGCLQSSLLIISVFLPGCAQNS 120	
Db	330 LGNSVDVMTSKVTAFNVETASPTSTPYDYVE----- 361	

Query Match	36.0%	Score 480.5	DB 20	Length 382
Best Local Similarity	62.6%	Pred No. 1	6e-38	

  

Query Match	36.0%	Score 480.5	DB 20	Length 382
Best Local Similarity	62.6%	Pred No. 1	6e-38	

Matches	90;	Conservative	2;	Mismatches	0;	Indels	0;	caps
Matches	90;	Conservative	2;	Mismatches	0;	Indels	0;	caps

Matches	89;	Conservative	2,	Information	1

QY 2 SVEEYNNKSAQNSTIPLIKITNTSDSLNDVKKVRYIYTSGTGGQGFMCDDHAGALL 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1 SVEEYNNKSAQNSTIPLIKITNTSDSLNDVKKVRYIYTSGTGGQGFMCDDHAGALL 60

OY 62 GNSYVDNNTSKVTANFVKETASPTSTYDYL 92  
 DB 61 GNSYVDNNTSKVTANFVKETASPTSTYDYL 91

# RESULT 14

AAW90080 ID AAW90080 standard; Protein: 156 AA.

XX AAW90080;

XX 09-MAR-1999 (first entry)

XX C. cellulovorans CBD-HSP fusion protein.

XX Cellulose binding domain; CBD; Cbpa; fusion protein; diagnosis; HSP;  
 KM Immunoblot; heat-shock protein; cross reactive protein; detection;  
 KM antigenic fragment; antibody; insulin-dependent diabetes mellitus;  
 KM cellulose; chitin; cellulolytic; amorphogenic.

XX Clostridium cellulovorans.  
 OS Synthetic.

XX US5856201-A.

XX 05-JAN-1999.

XX 27-OCT-1994; 94US-0330394.

XX 27-OCT-1994; 94US-0330394.

XX 14-APR-1993; 93US-0048164.

XX 14-APR-1994; 94WO-US04132.

XX (REGC ) UNIV CALIFORNIA.

XX (YISS ) YISSOM RES & DEV CO.

XX D01 RH, Goldstein MA, Shoseyov O, Shpiegl I;

XX WPI; 1999-105130/09.

XX Detection of a specific analyte by reaction with binding agent fused  
 PT to cellulose binding domain - and subsequent treatment with  
 PT cellulose and reaction of insoluble product with a label specific  
 PT for the analyte

XX Claim 4; Column 51-52; 63pp; English.

XX This sequence represents a fusion protein composed of the Cbpa protein  
 CC cellulose binding domain (CBD) from Clostridium cellulovorans and a heat  
 CC shock protein (HSP) fragment. This protein is used in diagnostic  
 CC immunoassays, e.g. to detect heat-shock proteins (HSP) and their  
 CC cross-reactive proteins, antigenic fragments or HSP-specific antibodies  
 CC (which indicate insulin-dependent diabetes mellitus, or susceptibility  
 CC to it. The CBD has very high affinity for cellulose (including  
 CC crystalline forms) and chitin (dissociation constant 0.8-1.4 mu M), but  
 CC has almost no cellulolytic or amorphogenic activities. The CBD binds over  
 CC a wide pH range and is not released from cellulose by washing with water.

XX Sequence 156 AA;

Query Match 35.4%; Score 472; DB 20; Length 156;

Best Local Similarity 97.8%; Pred. No. 3e-38; 0; Indels 0; Gaps 0;

Matches 89; Conservative 2; Mismatches 0; Gaps 0;

OY 2 SVEFYNSKSAQNTSIPPIIKITNTSDSLNDLVKRYRYTSDGQGFQFCHAGAL 61

DB 1 SVEFYNSKSAQNTSIPPIIKITNTSDSLNDLVKRYRYTSDGQGFQFCHAGAL 60

OY 62 GNSYVDNNTSKVTANFVKETASPTSTYDYL 92

DB 61 GNSYVDNNTSKVTANFVKETASPTSTYDYL 91

RESULT 15  
 AAR95080 ID AAR95080 standard; peptide: 167 AA.

XX AAR95080;

XX 30-OCT-1996 (first entry)

XX Cellulose binding domain.

XX Cellulose binding domain; CBD; hapten; moiety; biotin; avidin;

KM streptavidin; affinity chromatography; cell separation,

KM cell immobilisation; protein immobilisation; enzyme immobilisation;

KM multi-enzyme reactors; signal immunoassays; drug delivery; pesticide;

KM cellulose; chitin.

XX Clostridium thermocellum.

XX Key Location/Qualifiers

XX Misc-difference 139 /note= "Unidentified amino acid."

XX WO9613524-A1.

XX 09-MAY-1996.

XX 26-OCT-1995; 95WO-US13813.

XX 27-OCT-1994; 94IL-0111415.

XX (TECR ) TECHNION RES & DEV FOUND LTD.

XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

XX (YEDA ) YEDA RES & DEV CO LTD.

XX Bayer EA, Lamed R, Morag E, Shoham Y, Wilchek M;

XX WPI; 1996-239453/24.

XX Modified cellulose-binding domain mols. - having attached hapten,

PT partic. biotin, to provide an affinity system for, e.g. sepsis,

PT assays, reactors, delivery etc.

XX Claim 12; Page 34-35; 53pp; English.

XX A modified cellulose binding domain (CBD) or fraction of it, linked

CC to a hapten moiety via one or more Cys or Lys residues and complexes

CC comprising the biotinylated CBD and a biotin-binding molecule

CC selected from modified or unmodified avidin or streptavidin or an

CC anti-biotin antibody, can be used in affinity chromatography, cell

CC separation, cell, protein and enzyme immobilisation, selective

CC removal of biological materials, multi-enzyme reactors, signal

CC immunoassays and drug (e.g. pesticide) delivery. The CBD molecule

CC can be modified with the hapten without affecting its high affinity

CC for cellulose, or chitin.

XX Sequence 167 AA;

Query Match 18.4%; Score 245; DB 17; Length 167;

Best Local Similarity 51.5%; Pred. No. 6e-16;

Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

OY 1 MSVEFYNSKSAQNTSIPPIIKITNTSDSLNDLVKRYRYTSDGQGFQFCHAGAL 60

DB 9 LKVEFYNSNPSPDTTNSINPORKVNTGSSAIDLSKLTLYRYTYDGGQKDFQFCHAGAL 67

OY 61 LCN--SYVDNNTSKVTANFVKETASPTSTYDYLPSHMGCLQ 101

DB 68 IGSNGSYNGCITSNWKGFYKMS--TNNADRYLEISFGTLE 109

Search completed: October 27, 2002, 11:09:12

Job time : 35.0449 secs

**THIS PAGE BLANK (USPTO)**



## RESULT 2

054381 PRELIMINARY; PRT: 148 AA.  
 ID 054381  
 AC 054381;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE 17 KDA COMMON-ANTIGEN (FRAGMENT).  
 OS Rickettsia sp.  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=789;

SEQUENCE FROM N.A.  
 MEDLINE=98087556; PubMed=9425244;  
 RA Davis M.J., Ying Z., Brunner B.R., Ferwerda F.H.;  
 "Rickettsial relative associated with papaya bunchy top disease.";  
 RT Curr. Microbiol. 36:80-84(1998).  
 DR EMBL: U76907; AAC02809.1; -;  
 FT NON\_TER 1  
 SEQUENCE 148 AA; 15050 MW; A7AFEPDE0AE4C CRC64;

Query Match 22.7%; Score 303.5; DB 2; Length 148;  
 Best Local Similarity 40.1%; Pred. No. 2, 2e-17;  
 Matches 55; Conservative 30; Mismatches 43; Indels 9; Gaps 2;

OY 120 SREVGATGAVGAGVAGOLFEGSGRVSMAIGAVLGGLGSKIGSQMDQDK---IK 175  
 Db 17 NKQSGTILGTLGGLVGSQFGGTRLAAGAGALLIGNQAGMDQDKLAELT 76  
 OY 176 LNSLEKAVAGVTRNRNPDTGNSVPEPTRYKRYKOEEROQYCRFPOOKAMIAQOKO 235  
 Db 77 SORALEAPSGSSVWRNPDNGNYGTVPYSKAY-----KNTNGCYCRBYTQTVVVGKQO 131  
 OY 236 EITGTACPOPDGRNOVI 252  
 Db 132 KAYGTACROPDGOQOVV 148

## RESULT 3

09P9F2 PRELIMINARY; PRT: 159 AA.  
 ID 09P9F2  
 AC 09P9F2;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE 17 KDA GENUS-COMMON ANTIGEN.  
 OS Rickettsia felis (Rickettsia azadi).  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=42862;

SEQUENCE FROM N.A.  
 MEDLINE=21217364; PubMed=11321078;  
 RA Bouyer D.H., Stenos J., Crocquet-Valdes P., Moron C.G., Popov V.L.,  
 Zavala-Velazquez J.E., Foll L.D., Stohard D.R., Azad A.F.,  
 Walker D.H.;  
 "Rickettsia felis: molecular characterization of a new member of the  
 spotted fever group.";  
 RT Int. J. Syst. Evol. Microbiol. 51:339-347(2001).  
 DR EMBL: A195118; AAC28452.1; -;  
 SEQUENCE 159 AA; 16497 MW; 3AC5B020AF470A1F CRC64;

Query Match 21.2%; Score 283; DB 2; Length 159;  
 Best Local Similarity 37.9%; Pred. No. 1, 1e-15;  
 Matches 61; Conservative 30; Mismatches 54; Indels 16; Gaps 6;

OY 103 SLLIISV---FLVGC--AQNFSRQVGAATGAVGAGOLFEGSGRVSMAIG-GAVL 156  
 Db 5 SKMIITLASMLOACNGPCGMKQGTGTLGAGAGALLGSGQFGKQOL-VGVGVALL 63

OY 157 GGLIGSKIGSQMDQDK-----IKLNSLEKAVAGVTRNRNPDTGNSVPEPTRYKRYN 212  
 Db 64 GAVLGQIGAGDEDDRLAELTSGRALEATPSSVEMRPNDNGNYTPKNTY----- 119

OY 213 KQERROQYCRFPOOKAMIAQOKOITGTACPOPDGRNOVIS 253  
 Db 120 -RNSTGQYCRBYTQTVVVGKQOKAYGNACROPDGLMQOVN 159

## RESULT 4

052252 PRELIMINARY; PRT: 137 AA.  
 ID 052252  
 AC 052252;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE 17 KDA ANTIGEN (FRAGMENT).  
 OS Rickettsia cooley.  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=69410;

SEQUENCE FROM N.A.  
 RA Billings A.N., Telow G.J., Walker D.H.;  
 "Molecular characterization of a novel spotted fever group rickettsial  
 species from ixodes scapularis in Texas.";  
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF031534; AAB95267.1; -;  
 FT NON\_TER 1  
 SEQUENCE 137 AA; 14215 MW; A27597A9AFD85FC3 CRC64;

Query Match 19.4%; Score 258.5; DB 2; Length 137;  
 Best Local Similarity 39.3%; Pred. No. 9, 4e-14;  
 Matches 53; Conservative 25; Mismatches 46; Indels 11; Gaps 4;  
 OY 117 QNFSRQVGAATGAVGAGOLFEGSGRVSMAIG-GAVLGGLGSKIGSQMDQDK-- 173  
 Db 7 RGNKQGTILGAGAGALLGSGQFGKQOL-VGVGVALLGAVLGQIGAGMDQDKRL 65  
 OY 174 -IKLNSLEKAVAGVTRNRNPDTGNSVPEPTRYKRYKOEEROQYCRFPOOKAMIA 231  
 Db 66 AELTSGRALEAPSGSSVWRNPDNGNYGTVPKNTY-----RNSTGQYCRBYTQTVVIG 120

OY 232 GOKOITGTACPOPD 246  
 Db 121 GKQOKAYGNACROPD 135

## RESULT 5

09K2N6 PRELIMINARY; PRT: 144 AA.  
 ID 09K2N6  
 AC 09K2N6;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE 17KDA ANTIGEN (17 KDA ANTIGEN) (FRAGMENT).  
 OS male-killing Rickettsia from Adalia bipunctata.  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 Rickettsiaceae; Rickettsiae; Rickettsia.  
 NCBI\_TaxID=38028;

SEQUENCE FROM N.A.  
 RA Bernhard D., Hurst G.D., Majerus M.E.N.;  
 "On the evolution of male-killing Rickettsia (a-Proteobacteria) from two con-  
 transfer of male-killing Rickettsia (a-Proteobacteria) from two con-  
 generic ladybirds, Adalia bipunctata L. and A. decempunctata L.  
 (Coleoptera: Coccinellidae).";  
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ269518; CAB96383.1; -;  
 DR EMBL: AJ269517; CAB96382.1; -;  
 FT NON\_TER 1  
 SEQUENCE 144 AA; 144

```

DE 17 KDA ANTIGEN (FRAGMENT).
OS male-killing Rickettsia from Adalia decempunctata.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=120393;
RN [1]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX MEDLINE=20575219; PubMed=11133455;
RA Schlenker H.J.G.V.D., Habig M., Sloggett J.J., Webberley M.K.,
RA Bertrand D., Hurst G.D.D., Maternus M.E.N.;
RT "Incidence of male-killing Rickettsia spp. (alpha-Proteobacteria) in
RT the Ten-Spot Ladybird Beetle Adalia decempunctata L. (Coleoptera:
RT Coccinellidae).";
RL Appl. Environ. Microbiol. 67:270-277(2001).
DR EMBL; AJ269516; CAB96381.1; -.
FT NON TER 1 1
FT NON TER 144 144
SQ SEQUENCE 144 AA; 14801 MW; C825472F16A56AE7 CRC64;

Query Match 18.8%; Score 251.5; DB 2; Length 144;
Best Local Similarity 39.4%; Pred. No. 3.7e-13;
Matches 52; Conservative 24; Mismatches 45; Indels 11; Gaps 4;

OY 120 SRQEPGATGCAVGVCAQLREKSGGRSMATIG-GAVILGLIGSKTIGSMQDQK-----I 174
Db 17 NKQGGTLLKGAGGAGGALLSQFGKRGOL-VGVGVALLGAVIGGIGAGMDEQDRRLAEL 75
OY 175 KLNQSLLEKYGQGVVRWNPPTGNSGYSEVPVETCYGRYNKQERROOYCFEFOOKAMIAQOK 234
Db 76 TSQRALLEAPSSSNVEMRNPDNGNNGYTPNKTY-----RNSTGYCRHYTOTVYIGGKQ 130
OY 235 QEIVGTACPPQD 246
Db 131 QKSYGNACRQPD 142

RESULT 8
ID 045996 PRELIMINARY; PRT; 1546 AA.
AC 045996;
DT 01-NOV-1996 (TREMBlrel. 01. Created)
DT 01-MAR-2001 (TREMBlrel. 16. Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17. Last annotation update)
DE SCAFFOLDING PROTEIN PRECURSOR.
GN CIPC.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RN REVERBEL-LEROY C., TARDIF C., REVERBEL-LEROY C., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=96218696; PubMed=8636029;
RA Pages S., Belaïch A., Tardif C., Reverbel-Leroy C., Gaudin C.,
RA Belaïch J.P.;
RT "Interaction between the endoglucanase Cella and the scaffolding
RT protein CIPC of the Clostridium cellulolyticum cellulosome.";
RL J. Bacteriol. 178:2279-2286(1996).
RN [2]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [3]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [4]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [5]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [6]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [7]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [8]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [9]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [10]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [11]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [12]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [13]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [14]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [15]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [16]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
RA Belaïch J.P.;
RT "Sequence analysis of scaffolding protein CIPC and ORFXP, a new
RT cohesin-containing protein in Clostridium cellulolyticum: comparison
RT of various cohesin domains and subcellular localization of ORFXP.";
RL J. Bacteriol. 181:1801-1810(1999).
RN [17]
RN REVERBEL-LEROY C., TARDIF C., BELAÏCH A., BERNADAC A., GAUDIN C.,
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 35319; MEDLINE=10074072; PubMed=99173902;
RA Pages S., Belaïch A., Flerobe H.P., Tardif C., Gaudin C.,
```

```

RT      *Molecular study and overexpression of the Clostridium cellulolyticum
RT      celCCE cellulase gene in Escherichia coli."
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 35319;
RA      Pages S., Belatch A., Reverbel C., Tardif C., Fierobe H.P., Gaudin C.,
RA      Belatch J.P.;
RA      Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL: U40345; AAC28899.2; -.
DR      HSSP: Q06851; INBC.
DR      Interpro: IPR001936; CBD_3.
DR      Interpro: IPR002102; Cohesin.
DR      Interpro: IPR003880; Phosphopant_attach.
DR      Pfam: PF00942; CBD_3; 1.
DR      Pfam: PF00963; Cohesin; 8.
DR      Prodom: PD001947; CBD_3; 1.
DR      Prosite: PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.
KW      Signal.
FT      SIGNAL
FT      CHAIN
FT      SEQUENCE
SQ      1546 AA; 158748 MW; F8651504EC27809F CRC64;

Query Match
Best Local Similarity 18.7%; Score 249; DB 2; Length 1546;
Matches 72; Conservative 32; Mismatches 81; Indels 28; Gaps 7;

QY      1 MSVEFYNSNKAQNSTPIPIKINTSPDSDLNLDVRYRYTSDTQGTQGFMCDDHAGAL 60
QY      33 VSVGFNNGSSPAASSNSITARKRYNTSGSPINLADLKRYRYTQADAKPLTFMCDHAGYM 92
DB      61 LGNSVDNTSKVTANFEKETASPTSTDTYD---PSIMRCLOGSSLIITVFLVCAQ 117
DB      93 GSGNVIDATSKVTSF--KAVSPAYTNADHYLEVALNSDAGSLPAGGSTIEQTFRARNDWS 151
QY      118 NPSRVEAAGAVGVAGVAGOLFSGS---GRVSMAGVAVLGLSGSKIGSMDQDQ 173
DB      152 NFDOSNDMSYTA-----GSYMDMKISAFVGLAYG--STPGGNNPPQDP 197
QY      174 IKLNSLEKVRAG---QVTRMRNPTGTNSYS 201
DB      198 -TINPTISAKAGSPADTKITLTPTNGNTFNGIS 229

RESULT 9
QY      053154 PRELIMINARY; PRT; 154 AA.
AC      053154;
DT      01-NOV-1996 (Tremblrel. 01, Created)
DT      01-NOV-1996 (Tremblrel. 01, last sequence update)
DT      01-NOV-1996 (Tremblrel. 08, last annotation update)
DE      (CLONE PRB F15F 1), 5' END CDS (FRAGMENT).
OS      Rickettsia sp.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiellae; Rickettsia.
OX      NCBI_TaxID=789;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=93084757; PubMed=1452660;
RA      Bald R.W., Lloyd M., Stenos J., Ross B.C., Stewart R.S., Dwyer B.;
RA      "Characterization and comparison of Australian human spotted fever
RT      group rickettsiae.";
RT      J Clin. Microbiol. 30:2896-2902(1992).
DR      EMBL: M9391; AAA73386.1; -.
DR      NON_TER
SQ      SEQUENCE 154 AA; 15849 MW; F5C35855EDB439D2 CRC64;

Query Match
Best Local Similarity 18.3%; Score 244; DB 2; Length 154;
Matches 55; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

QY      103 SSIIITSV--FLVGC--AONFSRQEVGAATGAVGVAGOLFSGSGRVSMAG-GAVL 156

```

```

DB      5 SKIMITALATSLAQNCNGPGMNKOGTGLTGAGAGALLGSQFGKGGOL-VGVGVALL 63
QY      157 GGLISKIGGSMDOOK-----IKLNSLEKVRKAGVTFWRNPDGNSYSVEPRTYORYN 212
DB      64 GAVLGGQIVAGMBODRRLAELTISORALEAPSGSNVEMRNPNDGNYGVYTPPKTYRST 123
QY      213 KOERROQYCREFOQKAMIAQKOEIYGTACPPDGR 248
DB      124 GGD-----CRVYTVIVIGKQKQKAVGNACRPDQ 154

RESULT 10
QY      09F909 PRELIMINARY; PRT; 151 AA.
AC      09F909;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-MAR-2001 (Tremblrel. 16, last annotation update)
DE      OUTER MEMBRANE PROTEIN (FRAGMENT).
OS      Rickettsia helvetica.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiellae; Rickettsia.
OX      NCBI_TaxID=35789;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Nilsson K., Pahlson C.;
RT      "Novel peptide diagnostic reagent and kit for detection of
RT      rickettsiosis.";
RL      Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF181036; AAG09427.1; -.
FT      NON_TER
SQ      SEQUENCE 151 AA; 15621 MW; B77407B9C71E4B39 CRC64;

Query Match
Best Local Similarity 17.9%; Score 239; DB 2; Length 151;
Matches 55; Conservative 28; Mismatches 54; Indels 16; Gaps 6;

QY      103 SSIIITSV--FLVGC--AONFSRQEVGAATGAVGVAGOLFSGSGRVSMAG-GAVL 156
DB      5 SKIMITALATSLAQNCNGPGMNKOGTGLTGAGAGALLGSQFGKGGOL-VGVGVALL 63
QY      157 GGLISKIGGSMDOOK-----IKLNSLEKVRKAGVTFWRNPDGNSYSVEPRTYORYN 212
DB      64 GAVLGGQIVAGMBODRRLAELTISORALEAPSGSNVEMRNPNDGNYGVYTPPKTYRST 119
QY      213 KOERROQYCREFOQKAMIAQKOEIYGTACPP 245
DB      120 -RNSTGYCREYGTIVVIGKQKQKAVGNACROP 151

RESULT 11
QY      09F001 PRELIMINARY; PRT; 131 AA.
AC      09F001;
DT      01-MAR-2001 (Tremblrel. 16, Created)
DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-MAR-2001 (Tremblrel. 16, last annotation update)
DE      17 KDA PROTEIN (FRAGMENT).
OS      Rickettsia sp. California 2.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiellae; Rickettsia.
OX      NCBI_TaxID=147259;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CALIFORNIA 2;
RA      Raoult D.;
RA      "A new spf rickettsia isolated from fleas.";
RT      Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CALIFORNIA 2;
RA      Roux V., Raoult D.;
RA      Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

```



DR	EMBL, AF210693; AAG48554.1; -.
FT	NON_TER 1 1
FT	NON_TER 131 131
SO	SEQUENCE 131 AA; 13374 MW; 23C8819B29FFFB60 CRC64

Query Match	17.7%	Score 236.5	DB 2	Length 131
Best Local Similarly	38.3%	Pred. No. 5.6e12		
Matches 49: Conservative	24	Mismatches 44	Indels 11	Gaps 4

0y	120	SROEYGAATGAVNGVAGVGOLEKSGSGRMAIG -GALGGLISGKISQSDOODK-----I	174
0b	10	NGKGTGTLGGAGALLGSGPGKKGGL -GVGVGALLGAVLGGLGACAGDEDDRLAEL	68
0y	175	KLNSLEKVKAGQVTRMNPDTGNSYSVEPVRYQRYNKKORROQYCFEFOQKAMLAGQK	234
0b	69	TSQALALETPSGTSEVNRNPDPNGNHGVTPRYKT-----NSTSGQKREYTORVAVIGKK	123

QY	235	QEITYGTAC	242
		1:1111	
Db	124	QKAYGNAC	131

RESULT 12  
Q52637  
ID Q52637 PRELIMINARY; PRT; 131 AA.

DT	01-NOV-1996	(TREMblrel, 01, Created)
DT	01-NOV-1996	(TREMblrel, 01, last sequence update)
DT	01-NOV-1998	(TREMblrel, 08, last annotation update)

05 Rickettsia sp.  
0C Bacteria; Proteobacteria; alpha subdivision; Rickettsiales  
0C Rickettsiaceae; Rickettsiae; Rickettsia.

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94117373; Pubmed=8288533;  
PA WATSON T H, WATSON C D, ZHANG W

RA. *Majerus* M.E.;

RT. "Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*).";

RT. *Bacterio* 176:388-394(1994)

DR	EMBL:	U04102;	AAAI9235.1;	-
FT	NON_TER	1	1	
FT	NON_TER	131	131	
SO	SEQUENCE	131 AA:	13344 MW:	A1DCF71050DF52DF CRC64:

Query Match	17.78;	Score 236.5;	DB 2;	Length 131;
Best Local Similarity	38.38;	Pred. No. 5.6e-12;		
Matches 49; Conservative	24;	Mismatches 44;	Indels 11;	Gaps 4

QY	120	SRQEGAAATGAIVVGGVAGOLFCKGSGRYVSMaIG- GAVLGLGIGSKIGQSMQDQDK	---	1	174
Db	10	NKQCTGTLGGAGGALLISQFGKKGQOL- VGVGAGALLGAVLGGQISAGMDEQRIIAEL	----	1	68
QY	175	KLNSLSEKVKAGQVTRWENPDDTGSYSVEPVRTYQRYNKKDERQOYCREFOOKMAMIGQK	----	1	234
Db	69	TSQRRLLEAAPGSGNVENPNPDNGHGYTTPKKT- -----RNSTGQYCREYQTQVIVGSKQ	-----	1	123

QY	235	QEITYGTAC	242
		1:1111	
Db	124	QKAYGNAC	131

RESULT 13	
Q9L522	
ID Q9L522	PRELIMINARY;
Q9L522	PRT; 131 AA

DT 01-OCT-2000 (TREMBLEL. 15, Created)  
DT 01-OCT-2000 (TREMBLEL. 15, Last sequence update)  
DT 17-DEC-2001 (TREMBLEL. 19, Last annotation update)  
DE 17 KDA SURFACE ANTIGEN (FRAGMENT).  
OS Rickettsia peacockii.

0C Bacteria; Proteobacteria; alpha subdivision; Rickettsiales  
0C Rickettsiaceae; Rickettsiae; Rickettsia.  
0X NCBI\_TaxID=47589;

RP SEQUENCE FROM N.A.  
RC STRAIN=DAE100R;  
RX MEDLINE=21091941; PubMed=11157215;

RT "Isolation of a spotted fever group rickettsia, *Rickettsia peacockii*  
RT in a Rocky Mountain wood tick, *Dermacentor andersoni*, cell line.,"  
RL Appl. Environ. Microbiol. 67:546-552(2001).

FT	NON_TER	1	1
FT	NON_TER	131	131
SO	SEQUENCE	131 AA;	13413 MW; 228C020550CA9D0 CRC64;

Query Match	17.6%	Score 235.5	DB 2	Length 131
Best Local Similarity	38.3%	Pred. No. 6	7e-12	
Matches 49; Conservative	24;	Mismatches	44;	Indels 11; Gaps 4.

[illegible]

QY	235	QEIYGTAC	242
		:	
Db	124	QKAYGNAC	131

RESULT 14	
082830	
ID 082830	PRELIMINARY;
082830.	PRT; 1162 AA

DT	01-NOV-1998	(TREMblrel. 08, Created)
DT	01-NOV-1998	(TREMblrel. 08, Last sequence update)
DT	01-JUN-2001	(TREMblrel. 17, Last annotation update)
DE	SCAFOLDING PROTEIN	

GN CIPA.  
OS Clostridium josui.  
OC Bacteria; Firmicutes; Bacillus/clostridium group; Clostridiaceae.  
OC Clostridium.

OA	NCBI_1a1d-1455;
RN	[1]
RP	SEQUENCE FROM N.A
RX	MEDLINE=98361925;

RT scaffolding protein Cipa and identification of  
RT Cloning and DNA sequencing of the genes encoding Clostridium jostii  
RA Karita S., Sakka K., Ohmura K.;

RL J. Bacteriol. 180:4303-4308(1998).  
DR EMBL; AB004845; BAA32429.1; -.  
DR HSSP; Q06851; INBC.

DR InterPro; IPR002102; Cohesin.  
DR InterPro; IPR003880; Phosphopant\_attach  
DR Pfam; PF00942; CBD\_3; 1.

DR PRODOM; PD001947; CBD\_3; 1.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_5.  
SQ SEQUENCE 1162 AA; 120229 MW; 95F02DDDD27ADFBA2 CRC64;

Query Match	17.5%;	Score 233.5;	DB 2;	Length 1162;
Best Local Similarity	34.0%;	Pred. No. 1.6e-10;		
Matches	66;	Conservative	31;	Mismatches 74;
				Indels 23;
				Gaps 7

```

0y      1 MSVEFNSNKSQATNSINPIIKITNTSDSLNLNDVAKRYYYNSDGTGCGTFWCHAGAL 60
      ::::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     33 ISVQFNNGSSPISSSSIIARFKVNTSGSPINLADLKIRYYFTQDENKQMTFWCHAGYL 92

```

RESULT 15	
Q977Y4	
ID Q977Y4	PRELIMINARY;
	PRT; 1483 AA.

AC 01-OCT-2001 (TREMBLrel\_18, Created)  
 DT 01-OCT-2001 (TREMBLrel\_18, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel\_19, Last annotation update)  
 DE PROBABLY CELLULOSONAL CASHFOLDING PROTEIN, SECRETED,  
 DE CELLULOSE-BINDING AND COHESIN DOMAIN.  
 GN CAC0910.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxId=1486;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=15359325; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R.L., Lee H.M., Dubois J., Oiu D., Hittl J., Wolf Y.I.,  
 RA Tatusov R.L., Sabath F., Doucette-Stamm L., Soucaille P., Daly M.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.:  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum.";  
 RT J. Bacteriol. 183:4823-4838(2001).  
 RL EMBL: AE007606; MAK78886.1; -  
 DR InterPro: IPR001956; CBD\_3;  
 DR InterPro: IPR002102; Cohesin.  
 DR Pfam: PF00942; CBD\_3; 1.  
 DR Pfam: PF00963; Cohesin; 5.  
 DR ProDom: PD001947; CBD\_3; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1483 AA; 154444 MW; A7480E396AA9807 CRC64;

[illegible]

Search completed: October 27, 2002, 10:59:56  
Job time : 31.7392 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2002, 10:55:57 ; Search time 9.28497 Seconds  
(without alignments)  
1067.553 Million cell updates/sec

Title: US-09-677-374-6  
Sequence: 1 MSVEFYNSKNSAQTNSITPPL.....ITGTACPPQPGRMQVISTEK 256

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	1	CBPA_CLOCL
2	288	21.6	159	1	17KD_RICPR
3	286	21.4	159	1	17KD_RICJA
4	285	21.3	159	1	17KD_RICCN
5	276.5	20.7	159	1	17KD_RICCY
6	268	20.1	154	1	17KD_RICAU
7	259	19.4	154	1	17KD_RICPA
8	259	19.4	154	1	17KD_RICRH
9	255	19.2	154	1	17KD_RICMO
10	255	19.1	154	1	17KD_RICAM
11	245	18.4	772	1	CITPR_CLOTM
12	245	18.4	1853	1	CITPA_CLOTM
13	188.5	14.1	879	1	GUNT_CLOTM
14	188	14.1	986	1	GUNT_CLOSR
15	162.5	12.2	914	1	GUX2_CLOSR
16	156	11.7	80	1	17KD_RICCA
17	143.5	10.7	1039	1	GUNE_CALSA
18	143.5	10.7	1331	1	MANE_CALSA
19	138.5	10.4	1742	1	GUNA_CALSA
20	121	9.1	499	1	GUN2_BACSU
21	115	8.6	499	1	GUN3_BACSU
22	112.5	8.4	155	1	PCP_YEREN
23	111	8.3	145	1	YCEA_PAEIA
24	109.5	8.2	499	1	GUNI_BACSU
25	109	8.2	155	1	SLVB_SALTY
26	104.5	7.8	155	1	SLVB_ECOLI
27	103	7.7	504	1	GUNW_ERMCA
28	101.5	7.6	155	1	PCP_HAERY
29	101.5	7.6	505	1	GUNW_ERMCA
30	100.5	7.5	700	1	GUNA_PAEIA
31	100	7.5	700	1	YCF7_ECOLI
32	100	7.5	444	1	GUNW_ERMCA
33	98	7.3	922	1	PMPI_CHLPP

## ALIGNMENTS

RESULT 1	CBPA_CLOCL	STANDARD:	PRT: 1848 AA.
AC	P38058;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	Cellulose binding protein A precursor.		
GN	CBPA.		
OS	Clostridium cellulovorans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1493;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-92228810; PubMed-1565642;		
RA	Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;		
RT	"Primary sequence analysis of Clostridium cellulovorans cellulose		
RT	binding protein A.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).		
CC	FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE		
CC	ENZYMES.		
CC	-1- PTM: THE N-TERMINUS IS BLOCKED.		
CC	-1- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	EMBL: M73817; AAA23218.1; -.		
DR	PIR: A44140; A44140.		
DR	HSSP: Q06851; INBC.		
DR	InterPro: IPR001956; CBD_3.		
DR	InterPro: IPR002102; Cohesin.		
DR	Pfam: PF00942; CBD_3; 1.		
DR	Pfam: PF00963; Cohesin; 9.		
DR	ProDom: PD001947; CBD_3; 1.		
KW	Cellulose degradation; Cell wall; Repeat; Signal; glycoprotein.		
FT	SIGNAL. 29		
FT	CHAIN. 1 1848		
FT	DOMAIN. 29 189		
FT	SEQUENCE. 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;		
QY	Query Match	36.5%;	Score 487; DB 1; Length 1848;
QY	Best Local Similarity	46.8%;	Pred. No. 6.4e-32;
QY	Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;		
DB	1 MSVEFYNSKNSAQTNSITPPLIKITNTSDSLNLDVYVRYRYSDGOTGOTFCMDHAGAL 60		
DB	33 MSVEFYNSKNSAQTNSITPPLIKITNTSDSLNLDVYVRYRYSDGOTGOTFCMDHAGAL 92		
QY	61 LGNSYVNTSKSVTANFVKETASPTSTYDYLDPESHMGCGLQSSLLIISFLVCAQNFNS 120		

```

|||||
DB 93 LGNSYDNTSKVTANFVKETASPTSTYDYVE-----124
QY 121 ROEVGAATGAVGVAGOLFFGKSGRVSMAIGAVLGLIGSKIGOSMDQDKILNOSL 180
DB 125 -----FGFASGRATL-----KKGQFLTIQ-----143
QY 181 EKVAGQVTR--WRNPDTGNSYSVEPVRYORYNKQERROQCREFOQKAMTAGOKEIY 238
DB 144 -----GRITKSDMSNTYTQNDYSFSDASSSTPYVNP-----KVTGYIGGAK--VL 185

QY 239 GTACPOPD 246
DB 186 GTA-PGPD 192

RESULT 2
17KD_RICPR STANDARD; PRT; 159 AA.
ID 17KD_RICPR
AC P16624;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 17 kda surface antigen precursor.
GN OMP OR RP833.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA MEDLINE=89359171; PubMed=2768201;
RA Anderson B.E., Trianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: M28482; AAA26378.1; ALT_SEQ.
DB EMBL: AJ235273; CA15258.1; -.
DB PTR: D33971; D33971.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLICERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16672 MW; A33D404B65EB071 CRC64;
SQ
Query Match 21.6%; Score 288; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 7.2e-17;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
QY 103 SLLIISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFFGKSGRVSMAIG-CAVL 156

```

```

|||||
DB 5 SKMIIALASMLQACNGSGMKNKQGTLLGAGGALLGSGFGGKGQL-VGVGVGALL 63
QY 157 GGLIGSKIGOSMDQDK---IKNOSLEKVAGQVTRWRNPDTGNSYSVEPVRYORYN 212
DB 64 GAVLGGQIGASMDQDPRRLLETSQRALSSAPSNSIERNRNDNGHGYVTPNKTY-----119
QY 213 KOERROQYCREFOQKAMTAGOKEIYGTACPOPDGRMOVIS 253
DB 120 -RNSGQYCREYQTQTVILIGKQKQRTYGNACRQPDGMQVYN 159

RESULT 3
17KD_RICJA STANDARD; PRT; 159 AA.
ID 17KD_RICJA
AC Q52764;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 17 kda surface antigen precursor.
GN OMP.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA MEDLINE=95229950; PubMed=7714214;
RA Furuya Y., Katayama T., Yoshida Y., Kajino I.;
RT "Specific amplification of Rickettsia japonica DNA from clinical
RT specimens by PCR.";
RL J. Clin. Microbiol. 33:487-489(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: D16515; BAA03965.1; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLICERIDE (PROBABLE).
FT SEQUENCE 159 AA; 16554 MW; CDCE7CEBDCDB41 CRC64;
SQ
Query Match 21.4%; Score 286; DB 1; Length 159;
Best Local Similarity 37.9%; Pred. No. 1.1e-16;
Matches 61; Conservative 32; Mismatches 52; Indels 16; Gaps 6;
QY 103 SLLIISV---FLVGC--AQNFSROEVGAATGAVGVAGOLFFGKSGRVSMAIG-CAVL 156
DB 5 SKMIIALATMLQACNPGMKNKQGTLLGAGGALLGSGFGKGTGOL-VGVGVGALL 63
QY 157 GGLIGSKIGOSMDQDK---IKNOSLEKVAGQVTRWRNPDTGNSYSVEPVRYORYN 212
DB 64 GAVLGGQIGASMDQDPRRLLETSQRALSSAPSNSIERNRNDNGHGYVTPNKTY-----119
QY 213 KOERROQYCREFOQKAMTAGOKEIYGTACPOPDGRMOVIS 253
DB 120 -RNSGQYCREYQTQTVILIGKQKQRTYGNACRQPDGMQVYN 159

RESULT 4
17KD_RICCN STANDARD; PRT; 159 AA.
ID 17KD_RICCN
AC P05372;

```

```

DT 01-NOV-1988 (Rel. 09, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 17 kda surface antigen precursor.
GN OMP OR RC1287.
OS Rickettsia conorii, and
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision: Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxId=781, 783.
RN RN
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii, and R.rickettsii;
RX MEDLINE=89353171; PubMed=2768201;
RT Anderson B.E., Tzianabos T.;
RT "Comparative sequence analysis of a genus-common rickettsial antigen
RT gene.";
RL J. Bacteriol. 171:5199-5201(1989).
RN RN
RP SEQUENCE FROM N.A.
RC SPECIES=R.conorii; STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Andiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Reault D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RT Science 293:2093-2098(2001).
RN RN
RP SEQUENCE FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=8722152; PubMed=3108332;
RA Anderson B.E., Regnery R.L., Carlone G.M., Tzianabos T., McDade J.E.,
RA Fu Z.Y., Bellini W.J.;
RT "Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia
RT rickettsii.";
RL J. Bacteriol. 169:2385-2390(1987).
RN RN
RP SEQUENCE OF 1-30 FROM N.A.
RC SPECIES=R.rickettsii;
RX MEDLINE=89008059; PubMed=3139629;
RA Anderson B.E., Baumstark B.R., Bellini W.J.;
RT "Expression of the gene encoding the 17-kilodalton antigen from
RT Rickettsia rickettsii: transcription and posttranslational
RT modification.";
RL J. Bacteriol. 170:4493-4500(1988).
CC CC
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28479; AAA26379.1; -
DR EMBL: M28480; AAA26376.1; -
DR EMBL: AE008675; AAL03825.1; -
DR EMBL: M16486; AAA26381.1; -
DR EMBL: J03571; -; NOT_ANNOTATED_CDS.
DR PIR: A25972; A25972.
DR PIR: A31836; A31836.
DR PIR: A33971; A33971.
DR PIR: B33971; B33971.
DR PROSITE, PS00013; PROKAR_LIPOPROTEIN_1.
KW Outer membrane; Lipoprotein; Antigen; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 159 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT CONFLICT 146 146 N -> D (IN REF. 3).
FT CONFLICT 153 153 G -> E (IN REF. 3).
SQ SEQUENCE 159 AA; 16581 MW; 206A2BBF74FCE169 CRC64;

```

[illegible]

```

RESULT 6
17KD_RICAU STANDARD: PRT; 154 AA.
ID 17KD_RICAU
AC P50928;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia australis.
CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiae; Rickettsia.
CC NCBI_TaxID=787;
RN [1]
RP SEQUENCE FROM N.A.
RA Baird R.W., Ross B., Dwyer B.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M74042; AAA26394.1;
DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15967 MW; E3A833346FAC320 CRC64;

Query Match 20.1%; Score 268; DB 1; Length 154;
Best Local Similarity 37.8%; Pred. No. 3e-15;
Matches 59; Conservative 29; Mismatches 52; Indels 16; Gaps 6;

OY 103 SLLIIISV---FLVGAQ--NFSRQEVGAATGAVGVAGQLFGKSGGRVMAIG-GAVL 156
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 5 SKIMIALAASMLQACNSPGMKNKGCTGLLGAGGALGSGQKGGQL-VGVGALL 63
OY 157 GGLIGSKIGQSMDOOK-----IKLNOSLEKVKAGVTRMRNPDGNSYVEPVRYTORN 212
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 64 GAVLGQIGAGNDEDRRLAELTSORALETPASGSNVEWRPNNGNYVTYVTKYRN 123
OY 213 KOERROQYCRFEOQKAMTAGQKQETIYGTACPOPPGR 248
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 124 -----GQYCREYTGTVVIGGKQKAYGNACPOPDGQ 154

RESULT 7
17KD_RICPA STANDARD: PRT; 154 AA.
ID 17KD_RICPA
AC P50930;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia parkeri.
CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiae; Rickettsia.
CC NCBI_TaxID=35792;
RN [1]
RP SEQUENCE FROM N.A.
RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;
RA Pretzman C.I., Stothard D.R., Ralph D., Clark J.B., Fuerst P.A.;

```

```

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U1020; AAB07706.1;
DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154
SQ SEQUENCE 154 AA; 15897 MW; 5D06F45F9BDB5EEC CRC64;

Query Match 19.4%; Score 259; DB 1; Length 154;
Best Local Similarity 36.5%; Pred. No. 1.7e-14;
Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

OY 103 SLLIIISV---FLVGC--AQNFSRQEVGAATGAVGVAGQLFGKSGGRVMAIG-GAVL 156
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 5 SKIMIALATSLQACNSPGMKNKGCTGLLGAGGALGSGQKGGQL-VGVGALL 63
OY 157 GGLIGSKIGQSMDOOK-----IKLNOSLEKVKAGVTRMRNPDGNSYVEPVRYTORN 212
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 64 GAVLGQIGAGNDEDRRLAELTSORALETPASGSNVEWRPNNGNYVTYVTKYRN 119
OY 213 KOERROQYCRFEOQKAMTAGQKQETIYGTACPOPPGR 248
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 120 -RSTGQYCREYTGTVVIGGKQKAYGNACLOPDGQ 154

RESULT 8
17KD_RICRH STANDARD: PRT; 154 AA.
ID 17KD_RICRH
AC P50931;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 17 kDa surface antigen precursor (Fragment).
GN OMP.
OS Rickettsia rhipicephali.
CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiae; Rickettsia.
CC NCBI_TaxID=33992;
RN [1]
RP SEQUENCE FROM N.A.
RA Stothard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;
RA Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U1020; AAB07706.1;
DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Antigen; Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).
FT NON_TER 154 154

```

SEQUENCE 154 AA: 15895 MW: 0CF85AD5D96DEFFB CRC64:

Query Match 19.4%; Score 259; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 1.7e-14; Mismatches 53; Indels 16; Gaps 6;

Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 103 SSLIISV---FLVGC--AQNFSROEVGAATGAVGVAGQLRGKSGGRVMAIG-GAVL 156

DB 5 SKMITIALASMLQACNPGCGMKNKOGTGLLGAGGALLSGQGGKGGQ--VGVGVGALL 63

QY 157 GGLIGSKIGOSMDQDK---IKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYN 212

DB 64 GAVLGGOIGAGMDEODRRLAELTSQKALETPAGSGSNVEMRNPNGNNGYITPKKTY----- 119

QY 213 KOERROOYCREFOOKAMIAQOKOEITGTACPPQDGR 248

DB 120 -RNSTGOYCREYQTVVIGGKQOKKAYGNACLPDQ 154

RESULT 9

ID 17KD\_RICMO STANDARD: PRT: 154 AA.

AC P50929;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 kDa surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia montana.

OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_Taxid=33991;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OHIO 83-441;

RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U11017; AAB07705.1; -

DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19

FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).

FT NON\_TER 154 154

SEQUENCE 154 AA: 15881 MW: A09C53B8769E31DA CRC64:

Query Match 19.2%; Score 256; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 2.9e-14;

Matches 57; Conservative 30; Mismatches 53; Indels 16; Gaps 6;

QY 103 SSLIISV---FLVGC--AQNFSROEVGAATGAVGVAGQLRGKSGGRVMAIG-GAVL 156

DB 5 SKMITIALASMLQACNPGCGMKNKOGTGLLGAGGALLSGQGGKGGQ--VGVGVGALL 63

QY 157 GGLIGSKIGOSMDQDK---IKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYN 212

DB 64 GAVLGGOIGAGMDEODRRLAELTSQKALETPAGSGSNVEMRNPNGNNGYITPKKTY----- 119

QY 213 KOERROOYCREFOOKAMIAQOKOEITGTACPPQDGR 248

DB 120 -RNSTGOYCREYQTVVIGGKQOKKAYGNACLPDQ 154

RESULT 10

ID 17KD\_RICAM STANDARD: PRT: 154 AA.

AC P50927;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 17 kDa surface antigen precursor (Fragment).

GN OMP.

OS Rickettsia amblyomii.

OC Bacteria: Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_Taxid=33989;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MO 85-1084;

RA Stochard D.R., Ralph D.A., Clark J.B., Fuerst P.A., Pretzman C.;

RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U11013; AAB07704.1; -

DR PROSITE: PS00013; PROKAR.LIPOPROTEIN; 1.

KW Outer membrane; Lipoprotein; Antigen; Signal.

FT SIGNAL 1 19

FT CHAIN 20 >154 17 KDA SURFACE ANTIGEN.

FT LIPID 20 20 N-ACYL DIGLYCERIDE (PROBABLE).

FT NON\_TER 154 154

SEQUENCE 154 AA: 15879 MW: E4FB6429D943581 CRC64:

Query Match 19.1%; Score 255; DB 1; Length 154;

Best Local Similarity 36.5%; Pred. No. 3.5e-14;

Matches 57; Conservative 29; Mismatches 54; Indels 16; Gaps 6;

QY 103 SSLIISV---FLVGC--AQNFSROEVGAATGAVGVAGQLRGKSGGRVMAIG-GAVL 156

DB 5 SKMITIALASTLQACNPGCGMKNKOGTGLLGAGGALLSGQGGKGGQ--VGVGVGALL 63

QY 157 GGLIGSKIGOSMDQDK---IKLNOSLEKVKAGQVTRMRNPDTGNSYSVEPVRYQRYN 212

DB 64 GAVLGGOIGAGMDEODRRLAELTSQKALETPAGSGSNVEMRNPNGNNGYITPKKTY----- 119

QY 213 KOERROOYCREFOOKAMIAQOKOEITGTACPPQDGR 248

DB 120 -RNSTGOYCREYQTVVIGGKQOKKAYGNACLPDQ 154

RESULT 11

ID CIPB\_CLOTM STANDARD: PRT: 772 AA.

AC 001866;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SI/SL)

GN CIPB.

OS Clostridium thermocellum.

OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

OC Clostridium

OX NCBI\_Taxid=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS;

Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 CC Clostridium.  
 CC NCBI\_TaxID=1515;  
 RN (1)  
 RP SPOURCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.  
 RP STRAIN=ATCC 27405;  
 RX MEDLINE=93302508; PubMed=8316083;  
 RA Genecross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,  
 RA Demain A.L.;  
 RT "sequencing of a Clostridium thermocellum gene (cipa) encoding the  
 RT cellulosomal SL-protein reveals an unusual degree of internal  
 RT homology".  
 RL Mol. Microbiol. 8:325-334(1993).  
 RN (2)  
 RN SEQUENCE OF 1820-1853 FROM N.A.  
 RP MEDLINE=93209931; PubMed=8458833;  
 RX Fujino T., Beguin P., Aubert J.-P.,  
 RT "organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
 RT in attachment of the cellulosome to the cell surface".  
 RL J. Bacteriol. 175:1891-1899(1993).  
 RN (3)  
 RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.  
 RX MEDLINE=97238934; PubMed=9083107;  
 RA Shimon I.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,  
 RA Frolow F.;  
 RT "A cohesin domain from Clostridium thermocellum: the crystal  
 RT structure provides new insights into cellulosome assembly";  
 RL Structure 5:381-390(1997).  
 RN (4)  
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.  
 RX MEDLINE=97076134; PubMed=8918451;  
 RA Torro J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,  
 RA Steitz T.A.;  
 RT "Crystal structure of a bacterial family-III cellulose-binding  
 RT domain: a general mechanism for attachment to cellulose";  
 RL EMBO J. 15:5739-5751(1996).  
 RN (5)  
 RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.  
 RX MEDLINE=98022914; PubMed=9402065;  
 RA Tavares G.A., Beguin P., Alzari P.M.;  
 RT "The crystal structure of a type I cohesin domain at 1.7-A  
 RT resolution";  
 RL J. Mol. Biol. 273:701-713(1997).  
 CC -I- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT  
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE  
 CC CELLULOYTIC ENZYMES.  
 CC -I- SUBCELLULAR LOCATION: CELL SURFACE.  
 CC -I- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150  
 CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC  
 CC COMPONENTS OF THE CELLULOSE.  
 CC -I- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions in no way  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: L08665; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: X67506; CAA47840.1; -;  
 DR F036859; S36859.  
 DR PDB: 1AND; 23-JUL-97.  
 DR PDB: 1AOH; 08-JUL-98.  
 DR PDB: 1NBC; 26-SEP-97.  
 DR InterPro: IPR001956; CBD\_3.  
 DR InterPro: IPR002102; Cohesin.  
 DR InterPro: IPR002105; Dockerin\_1.  
 DR InterPro: IPR002048; EF-hand.  
 DR



DR Pfam: PF00942; CBD\_3; 1.  
 DR Pfam: PF00963; Cohesin; 9.  
 DR Pfam: PF00404; Dockerin; 1; 2.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN; 1.  
 DR PROSITE: PS00448; CLOS\_CELLOSOME\_RPT; 2.  
 DR Cellulose degradation: Cell wall; glycoprotein; Repeat; Signal;  
 KW 3D-structure.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1853  
 FT DOMAIN 364 522  
 FT DOMAIN 34 1694  
 FT REPEAT 29 182  
 FT REPEAT 183 322  
 FT REPEAT 560 704  
 FT REPEAT 724 866  
 FT REPEAT 889 1031  
 FT REPEAT 1054 1196  
 FT REPEAT 1219 1361  
 FT REPEAT 1384 1526  
 FT REPEAT 1548 1690  
 FT DOMAIN 1791 1846  
 FT REPEAT 1791 1846  
 FT REPEAT 1824 1847  
 FT REPEAT 1615 1615  
 FT CONFLICT 1615 1615  
 FT SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 18.4%; Score 245; DB 1; Length 1853;  
 Best Local Similarity 51.5%; Pred. No. 4; 1e-12;  
 Matches 53; Conservative 16; Mismatches 30; Indels 4; Gaps 3;

QY 1 MSYEFVNSKSAQTNSITPIKIRITNTSDSLNDVKKVRYRYSDTGQGTFCMDHAGAL 60  
 Db 369 LKVEFVNSKSAQTNSITPIKIRITNTSDSLNDVKKVRYRYSDTGQGTFCMDHAGAL 60  
 QY 61 LGN-STVDNTSKVTANFKETASPTSTYDYLDPSPHMRCLQ 101  
 Db 428 IGSNGSYNCITSNVKGFVKMSS-TNNADTYLEISFTGTL 469

RESULT 13  
 GUNZ\_CLOTM STANDARD; PRT; 879 AA.  
 ID GUNZ\_CLOTM  
 AC 002934;  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 32, Last annotation update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Endoglucanase I precursor (EC 3.2.1.4) (EGI) (Endo-1,4-beta-glucanase)  
 DE (Cellulase I).  
 OS Clostridium thermocellum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1515;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-69.  
 RC STRAIN-NCIB 10682;  
 RA MEDLINE-93171873; PubMed-8436949;  
 RA Hazlewood G.P., Davidson K., Laurie J.I., Huskisson N.S.,  
 RA Gilbert H.J.;  
 RT "Gene sequence and properties of Cell, a family E endoglucanase from  
 RT Clostridium thermocellum";  
 RL J. Gen. Microbiol. 139:307-316(1993).  
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOMYDOLYSIS OF 1,4-BETA-  
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL GLUCAN.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose.  
 CC -1- PATHWAY: CELLULOSE DEGRADATION.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL  
 CC HYDROLASES).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L04735; AAA20892.1; -  
 DR PIR: A47704; A47704.  
 DR HSP: P26221; 1TF4.  
 DR InterPro: IPR001936; CBD\_3.  
 DR InterPro: IPR001701; Glyco\_hydro\_9.  
 DR Pfam: PF00942; CBD\_3; 2.  
 DR Pfam: PF00759; Glyco\_hydro\_9; 1.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR PROSITE: PS00592; GLYCOSYL\_HYDROL\_P9\_1; 1.  
 DR PROSITE: PS00698; GLYCOSYL\_HYDROL\_P9\_2; 1.  
 DR Cellulose degradation; Hydrolyase; Glycosidase; Signal.  
 KW SIGNAL 1 55  
 FT CHAIN 56 879  
 FT DOMAIN 56 518  
 FT DOMAIN 729 879  
 FT ACT\_SITE 448 448  
 FT ACT\_SITE 486 486  
 FT ACT\_SITE 495 495  
 FT SEQUENCE 879 AA; 97796 MW; 35A60069A514A927 CRC64;

Query Match 14.1%; Score 188.5; DB 1; Length 879;  
 Best Local Similarity 28.4%; Pred. No. 7; 3e-08;  
 Matches 59; Conservative 20; Mismatches 48; Indels 81; Gaps 9;

QY 3 VEFVNSKSAQTNSITPIKIRITNTSDSLNDVKKVRYRYSDTGQGTFCMDHAGAL 62  
 Db 742 LQVANGAGATNSINPRKILNNGTAINLSVKKIRYYTKEGASQNFCDHSSA-G 799  
 QY 63 NSYVDNTSKVTANFKETASPTSTYDYLDPSPHMRCLQSSLIITVFGAQNFSRQ 122  
 Db 800 N-SNVTGNF-NLSPKEGADT-CL- 821  
 QY 123 EVGATGAVVGVAGQGFKGSGVSMAGCAVIGLIGSKIGSMQDQKIKINQSLK 182  
 Db 822 EVG-----FGSGAG--TLDPGGSV-----EVOIRFSKED 848  
 QY 183 VKAGQVTRMRNPDGNSYSVEPRTYR 210  
 Db 849 -----WSNYNSNDISFKQACLROR 868

RESULT 14  
 GUNZ\_CLOS STANDARD; PRT; 986 AA.  
 ID GUNZ\_CLOS  
 AC P23659;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Endoglucanase 2 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Thermocatalytic cellulase) (Avicelase I).  
 OS Clostridium stercorarium.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1510;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-36 AND 475-486.  
 RC STRAIN-NCIB 11745;  
 RA MEDLINE-91066838; PubMed-2250652;  
 RA Jauris S., Ruecknagel K.P., Schwarz W.H., Kratzsch P.,  
 RA Bronnenmeier K., Staudenbauer W.L.;  
 RT "Sequence analysis of the Clostridium stercorarium cel2 gene encoding  
 RT a thermocatalytic cellulase (Avicelase I): Identification of catalytic  
 RT and cellulose-binding domains";  
 RL Mol. Gen. Genet. 223:258-267(1990).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

```

RN      [12]
RP      CHARACTERIZATION.
RC      SMRAIN-NCIB 11754;
RX      MEDLINE=91364666; PubMed=1909625;
RA      Bronnenmeier K., Ruecknagel K.P., Staudenbauer W.L.:
RT      "Purification and properties of a novel type of
RT      exo-1,4-beta-glucanase (avicelase II) from the cellulolytic
RT      thermophile Clostridium stercoararium.",
RL      Eur. J. Biochem. 200:379-385(1991).
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC      in cellulose and cellobiose, releasing cellobiose from the non-
CC      reducing ends of the chains.
CC      -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
CC      HYDROLASSES).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation - its
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC      or send an email to license@isb-slb.ch).
CC      -----
DR      EMBL: Z69359; CAA93280.1; -.
DR      HSSP: Q06851; INBC.
DR      InterPro: IPR001956; CBD_3.
DR      InterPro: IPR000556; Glyco_hydro.48.
DR      Pfam: PF00942; CBD_3. 1.
DR      pfam: PF02011; Glyco_hydro.48; 1.
DR      PRINTS: PR00844; GLHYDRLASE48.
DR      PRODOM: PD001947; CBD_3; 1.
DR      ProDom: PD011903; Glyco_hydro.48; 1.
DR      ProDom: PD011903; Glyco_hydro.48; 1.
KW      Cellulose degradation; hydrolase; Glycosidase; signal.
FT      SIGNAL 1 33
FT      CHAIN 34 914
SQ      SEQUENCE 914 AA: 103020 MW; DDBB6017D6DF82C CRC64:
QY      Query Match 12.2%; Score 162.5; DB 1; Length 914;
QY      Best Local Similarity 37.8%; Pred. No. 1e-05;
QY      Matches 34; Conservative 17; Mismatches 30; Indels 9; Gaps 2
QY      3 VEFVNSNRKQQTNSITPIKITTSDSLNDVKKVYRTSTGCTGGCFWCDHAGALLG 62
QY      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      Db 769 IQSNANMNGEISIMPRFRIRYNGS:::|::|::|::|::|::|::|::|
QY      63 NSYVDNTSKVTANVFKETASPTSYDYRLD 92
QY      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      Db 826 -----GSSNVTGTFFVKMDGATGA-DYLE 849

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 27, 2002, 10:56:01 ; Search time 16.8014 Seconds

(without alignments)  
1464.098 Million cell updates/sec

Title: US-09-677-374-6

Perfect score: 1335  
Sequence: 1 MSVEFYNSKSAQNSITPPI.....IYGTACPPDGRHQVISTEK 256

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

PIR-71:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	36.5	1848	2 A44140	cellulose-binding
2	288	21.6	159	2 D33971	ricketsial common
3	285	21.3	159	2 B33971	ricketsial common
4	285	21.3	159	2 A33971	ricketsial common
5	285	21.3	159	2 G97860	17K surface antigen
6	276.5	20.7	159	2 C33971	ricketsial common
7	276	20.7	159	2 A25872	17K antigen precursor
8	249	18.7	586	2 PC6006	scaffolding protein
9	245	18.4	1854	2 S36859	clpA protein - Clp
10	233.5	17.5	1162	2 T30433	scaffolding protein
11	223.5	16.7	1483	2 C97012	probably cellulose
12	188.5	14.1	879	2 A47704	endoglucanase I (E
13	188	14.1	986	2 S12021	thermoactive cellu
14	151.5	11.3	1779	2 T31085	xylinase - Caldice
15	147	11.0	915	2 A43802	xylinase (EC 3.2.
16	143.5	10.7	1039	2 S02711	cellulase (EC 3.2.
17	143.5	10.7	1331	2 A48954	cellulase (EC 3.2.
18	142	10.6	1711	2 T31337	mannan endo-1,4-be
19	138.5	10.4	1742	2 T17120	1,4-beta-glucanase
20	127.5	9.6	182	2 D83169	cellulase (EC 3.2.
21	122	9.1	257	2 F87413	conserved hypochet
22	121	9.1	508	2 G69593	hypochet (EC 3.2.
23	120.5	9.0	131	2 A13418	cellulase (EC 3.2.
24	115	8.6	489	2 JN0111	17K surface antigen
25	113.5	8.5	499	2 A27198	cellulase (EC 3.2.
26	112.5	8.4	155	2 S23787	outer membrane lip
27	111.5	8.4	154	2 B83514	conserved hypochet
28	111	8.3	145	2 A41897	cellulase homolog
29	109.5	8.2	508	2 A26874	cellulase (EC 3.2.

30	109	8.2	155	2 A10693	outer membrane lip
31	109	8.2	179	2 AE0644	probable secreted
32	109	8.2	232	2 G87629	hypochet
33	109	8.2	257	2 B82837	conserved hypochet
34	107.5	8.1	142	2 AD2696	lipA protein (lipo
35	107.5	8.1	155	2 AF0289	probable lipoprote
36	107	8.0	125	2 D97478	lipA protein (lipo
37	107	8.0	37	2 I40548	bifunctional cellu
38	106.5	8.0	155	2 AG0443	outer membrane lip
39	104.5	7.8	155	2 C64921	outer membrane lip
40	104.5	7.8	155	2 F90922	probable outer mem
41	104.5	7.8	139	2 B85771	probable outer mem
42	104	7.8	135	2 S58234	lipA protein - Rhl
43	103	7.7	232	2 AD3350	outer membrane pro
44	103	7.7	504	2 S54744	cellulase (EC 3.2.
45	101.5	7.6	155	2 T64130	PAL cross-reacting

## ALIGNMENTS

RESULT 1  
A44140  
cellulose-binding protein A precursor - Clostridium cellulovorans  
C:Species: Clostridium cellulovorans  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 15-Oct-1999  
C:Accession: A44140  
R:Shoseyov, O.; Takagi, M.; Goldstein, M.A.; Doi, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3483-3487, 1992  
A:Title: Primary sequence analysis of Clostridium cellulovorans cellulose binding pro  
A:Reference number: A44140; MUID:92228810  
A:Accession: A44140  
A>Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-1848 <SHO>  
A:Cross-references: GB:M73817; NID:g144748; PIDN:AAA23218.1; PID:g144749  
C:Genetics:  
A:Gene: cbpA

Query Match 36.5%; Score 487; DB 2; Length 1848;  
Best Local Similarity 46.8%; Pred. No. 2.8e-32;  
Matches 116; Conservative 12; Mismatches 30; Indels 90; Gaps 7;

QY	1	MSVEFYNSKSAQNSITPPIIKITNTSDSLNDLVKRYRYSYSDGQGFQFCDHAGAL	60
DB	33	MSVEFYNSKSAQNSITPPIIKITNTSDSLNDLVKRYRYSYSDGQGFQFCDHAGAL	92
QY	61	LGNSYVNTSKVTANFYKETASPTSTYDYLDPSPHMRCLQSSLIITSVFLVCAGNFS	120
DB	93	LGNSYVNTSKVTANFYKETASPTSTYDYLDPSPHMRCLQSSLIITSVFLVCAGNFS	124
QY	121	RQEVGAATGAVGVAGQLFGKSGRVSMAIGAVLGLIGSKIGQSMDDQDKIKLNSL	180
DB	125	-----FGFASGRATL-----KKGFITID-----	143
QY	181	EKVAGQVTR--WRNPDTGNSYSEPVRTYORYNKKEROOYCEFOQKAMIAQOKOET	238
DB	144	-----GRTRSDMSNYQTNDYSPDASSTPVPV-----RVYTGIGAK--VL	185
QY	239	GTACPOP 246	
DB	186	CTA-PCPD 192	

RESULT 2  
D33971  
ricketsial common antigen precursor - Rickettsia prowazekii  
N:Alternate names: 17kD surface antigen; outer membrane protein (omp); RP833  
C:Species: Rickettsia prowazekii  
C:Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 03-Nov-2000  
C:Accession: D33971; B71645  
R:Anderson, B.E.; Tzianabos, T.  
J. Bacteriol. 171, 5199-5201, 1989

[illegible]

RESULT 5  
G97860  
17k surface antigen precursor [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C:Accession: G97860  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: G97860  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <KUR>  
A:Cross-references: GB:AE006914; PIDN:ALU03825.1; PID:915620425; GSPDB:GN00173  
C:Genetics:  
A:Gene: omp  
C:Superfamily: rickettsial common antigen

Query Match 285; DB 2; Length 159;  
Best Local Similarity 37.9%; Pred. No. 1.le-16;  
Matches 61; Conservative 31; Mismatches 53; Indels 16; Gaps 6;

QY	103	SSLIIITSV---FLVAC--AQNFSQDEVGATGAVVGVAGOLFEGSGSRVSMZIG--GAVL	159
		:     :     :     :     :     :     :     :     :     :     :	
Db	5	SKIMIIALTATSLMQACNPGSGMNRKQGTGTLTGAGAGALLGSPFGAKGKOL--VGVGVGALL	63
		:     :     :     :     :     :     :     :     :     :	
QY	157	GGLIGSKIGQSDMDQDK---IKLNQSLIEKVKAGQVTRWRNPDTGNSTSVEPARTYQRYN	212
		:     :     :     :     :     :     :     :     :     :	
Db	64	GAVVGGQICAGNDEBDRRLAELTQSRLALETAPSGSNVNEENPDNGNYGYVPNKTY----	119
		:     :     :     :     :     :     :     :     :     :	
QY	213	KQERRQDYCREFOQKAMTAGQKQELTYGTACPDQPDGRHMYVS	253
		:     :     :     :     :     :     :     :     :     :	
Db	120	-RNSTGQYCREYTGTVVIGGKQAKAYGNACRQPDQGMQVYN	159

Rickettsial common antigen precursor - Rickettsia typhi

C:Species: Rickettsia typhi

C>Date: 16-Mar-1990 #sequence\_revision 16-Mar-1990 #text\_change 20-Aug-1999

C:Accession: C33971

R:Anderson, B.E.; Tzianabos, T.

J. Bacteriol. 171, 5199-5201, 1989

A:Title: Comparative sequence analysis of a genus-common rickettsial antigen gene.

A:Reference number: A33971; MUID:89359171

A:Accession: C33971

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-159 <AND>

A:Cross-references: GB:M28481; NID:g152459; PIDN:AAA26377.1; PID:g152460

C:Superfamily: rickettsial common antigen

Query Match 20.7%; Score 276.5; DB 2; Length 159;

Best Local Similarity 39.6%; Pred. No. 5.8e-16;

Matches 55; Conservative 27; Mismatches 46; Indels 11; Gaps 4;

OY 120 SROEVGAATGAVGVAGOLFQKSGRVSMAIG-GAVLGGLIGSKIGSDMOODK---I 174

DB 27 NKQGTGLLGGAGGALGSGFGHCKGQL-VGVGALLGAVLGQICASLDEDDKLLER 85

OY 175 KLNQSLKVKAGVTRMRNPDTGNSVSEPVRYQRYNKOERROQYCRFQOKAMIAQOK 234

DB 86 TSQALSSAPSGSNIEMRNDNGHGVTPNKTY-----RNSTGQYCREYTVIGVIGKQ 140

OY 235 QEYRTGCPQPDGRMVOIS 253

DB 141 QTTYNACRQPDGMQVYN 159

#### RESULT 7

A25972 17K antigen precursor - Rickettsia rickettsii

C:Species: Rickettsia rickettsii

C>Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 20-Aug-1999

C:Accession: A25972

R:Anderson, B.E.; Regnery, R.L.; Carlone, G.M.; Tzianabos, T.; McBride, J.E.; Fu, Z.Y.; F

J. Bacteriol. 169, 2385-2390, 1987

A:Title: Sequence analysis of the 17-kilodalton-antigen gene from Rickettsia rickettsii.

A:Reference number: A25972; MUID:87222152

A:Accession: A25972

A:Molecule type: DNA

A:Residues: 1-159 <AND>

A:Cross-references: GB:M16486; NID:g152467; PIDN:AAA26381.1; PID:g152468

C:Superfamily: rickettsial common antigen

Query Match 20.7%; Score 276; DB 2; Length 159;

Best Local Similarity 37.3%; Pred. No. 6.4e-16;

Matches 60; Conservative 31; Mismatches 54; Indels 16; Gaps 6;

OY 103 SLLIISV---FLVGC--AQNFSHOEVAATGAVGVAGOLFQKSGRVSMAIG-GAVL 156

DB 5 SKIMIALATSMLOACNGPGMKNKGCTGLLGGAGALLGSGFGKQQL-VGVGALL 63

OY 157 GGLIGSIQSDMOODK---IKLNQSLKVKAGVTRMRNPDTGNSVSEPVRYQRYN 212

DB 64 GAVLGQIGAGMDQDRRLAELTSSQALETAPSGSNVEMRNPDNGNGYVTPNKTY---- 119

OY 213 KOERROQYCRFQOKAMIAQOKYEYTRACQPDGRMVOIS 253

DB 120 -RNSTGQYCREYTVIGVIGKQAKAYDQACQPDGMQVYN 159

#### RESULT 8

PC6006 scaffolding protein ClpC precursor - Clostridium cellulolyticum (fragment)

C:Species: Clostridium cellulolyticum

C>Date: 17-Aug-1990 #sequence\_revision 16-Aug-1996 #text\_change 13-Nov-1998

C:Accession: PC6006

R:Pages, S.; Belach, A.; Tardif, C.; Reverbel-Leroy, C.; Gaudin, C.; Belach, J.P.

J. Bacteriol. 178, 2279-2286, 1996

A:Title: Interaction between the endoglucanase CelA and the scaffolding protein ClpC

A:Reference number: PC6006; MUID:96218696

A:Accession: PC6006

A:Molecule type: DNA

A:Residues: 1-586 <PAG>

A:Cross-references: GB:U40345; NID:g1314291; PID:g1314292

A:Experimental source: ATCC 35319

C:Genetics:

A:Gene: clpC

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-586/Product: scaffolding protein ClpC #status predicted <MAT>

F:28-187/Domain: cellulose binding #status predicted <CBD>

F:188-292/Domain: hydrophilic <HYL>

F:293-439/Domain: hydrophobic <HYB1>

F:440-586/Domain: hydrophobic <HYB2>

Query Match 18.7%; Score 249; DB 2; Length 586;

Best Local Similarity 33.8%; Pred. No. 5.7e-13;

Matches 72; Conservative 32; Mismatches 81; Indels 28; Gaps 7;

OY 1 MSVEFYNSKSAQNSITPIKITNTSDSLNDVAVRYRTSDGTQGTFCMDHAGAL 60

DB 33 VSVQFNNQSSPSSNSIYAREKVTNTGSPINLADLRYYTYTDADKPLTFWCDHAGYM 92

OY 61 LGNSVYDNTSKVTANFKETASPTSTYDTYLD--PSHMRCCLOGSSLLIISVFLVCAQ 117

DB 93 SGSNTIDATSKVTGSF-KANSPAVTNMDHYLEVALNSDAGSLPGSGSEIOTRRARNDMS 151

OY 118 NFSROEVAATGAVGVAGOLFQKGS---GRVSMAIGAVLGGLIGSKIGSDMOODK 173

DB 152 NFSQSDMSVYTA-----GSYMDQKISAFVGTGLAVG--STPDGGRPPQDP 197

OY 174 IKLNQSLKVKAG-----QVTRMRNPDTGNSYS 201

DB 198 -TIMPTISAKAGSPADTKITLTPNGNTPNGIS 229

#### RESULT 9

S36859 ClpA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL

C:Species: Clostridium thermocellum

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000

C:Accession: S36859; S3527; S25767; S28659; T18261

R:Gerngross, U.T.; Demain, A.L.

Submitted to the EMBL data library, January 1993

A:Reference number: S36859

A:Accession: S36859

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1854 <GER>

A:Cross-references: EMBL:L08665

R:Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.

Mol. Microbiol. 8, 325-334, 1993

A:Title: Sequencing of a Clostridium thermocellum gene (clpA) encoding the cellulosome

A:Reference number: S33527; MUID:93302508

A:Accession: S33527

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1615,1617-1854 <GE2>

A:Cross-references: EMBL:L08665

R:Fujino, T.; Beguin, P.; Aubert, J.P.

FEMS Microbiol. Lett. 94, 165-170, 1992

A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides tha

A:Reference number: S25767

A:Accession: S25767

A:Molecule type: DNA

A:Residues: 1216-1237,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1854 <FUJ>

A:Cross-references: EMBL:X67406

R:Beugin, P.

submitted to the EMBL data library, August 1992

A:Reference number: S28659

A:Accession: S28659

A:Molecule type: DNA  
A:Residues: 1216-1232,'Y','Y',1234-1241,'T','1243-1319','R',1321-1615,1617-1768,'R',1770-1854  
A:Cross-References: EMBL:X67406  
R:Fujino,T.; Beguin,P.; Aubert,J.P.  
J.Bacteriol. 175, 1891-1899, 1993  
A>Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosome  
e.  
A:Reference number: Z18847; MUID:93209931  
A:Accession: T18261  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1821-1854 <F02>  
A:Cross-References: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAAT840.1  
C:Genetics:  
A:Gene: ctpA

[illegible]

```

RESULT 10
T30433
scaffolding protein - Clostridium josui
C:Species: Clostridium josui
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30433
R:Kakuchi, M.; Isui, A.; Suzuki, K.; Fujino, T.; Fujino, E.; Kimura, T.; Karita, S.; Sa-
U. Bacteriol. 180, 4303-4308, 1998
A>Title: Cloning and DNA sequencing of the genes encoding Clostridium josui scaffolding
e.
A:Reference number: Z20837; MUID:98361925
A:Accession: T30433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1162 <KKA>
A:Cross-references: EMBL:AB004845; NID:g3445476; PIDN:BA52429.1; PID:g3445477
C:Genetics:
A:Gene: cipa

```

[illegible]

RESULT 11  
C97012  
probably cellulosomal scaffolding protein precursor, secreted, cellulose-binding and cot  
C;Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: C97012  
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97012  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1483 <GR> PIDN:AAK78886.1; PID: g15023809; GSPDB:GN00168  
A:Cross-references: KB:AE001437; A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0910

[illegible]

QY 181 EKVAGQVTRRNPDTGNSYS 201  
| | | | |  
Db 154 -----WSNFDQSDNYS 164

RESULT 12

A47704

endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermocellum

C:Species: Clostridium thermocellum

C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999

C:Accession: A47704

R:Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Hushisson, N.S.; Gilbert, H.J.

J. Gen. Microbiol. 139, 307-316, 1993

A:Title: Gene sequence and properties of Cell, a family E endoglucanase from Clostridium

A:Reference number: A47704; MUID:93171873

A:Accession: A47704

A:Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-879 <HA>

A:Cross-references: GB:LO4735; NID:g144807; PIDN:AAA20892.1; PID:g144808

A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIIP:125638)

C:Keywords: glycosidase; hydrolase

```

Query Match          14 1%; Score 188.5; DB 2; Length 879;
Best Local Similarity 28.4%; Pred No. 1,1e-07;
Matches 59; Conservative 20; Mismatches 48; Indels 81; Gaps 9.

OY      3 VEFNRSKSAQTNSITPIIKIINTSDSLINDVAKRYRYYTSDGCGOTFPCDHAGALLG 62
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       742 LQYANGAGATSNINSINERFKIINNCKRAINLDVKIRYYTREGASQNFCDMSSA -G 799
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      63 NSYDNTSKYTANVKEFTASTPTDYTLDPSHRCLOGSSLIIISVELGCANFSRQ 122
        :|||:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       800 N-----SNTGNPF-NLSSPKRGADT-----CL-----821
        :|||:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY     123 EVGAATGAVGVGAVAGOLFPGKSGSRVSMATGAVLGILSKIGSQMDQDRKIKNQLSEK 182
        |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       822 EVG-----FGSGAG--TIDPQGSV-----EVOIRFSKED 848
        :|||:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY    183 VKACGVTFRMRNDPTGNSYSVEPEVRYOR 210
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 849 -----WSNYNSNDYSEFKOACLROR 868

## RESULT 13

T31021

thermoactive cellulase - Clostridium stercorarium

N:Alternate names: avicelase I

C:Species: Clostridium stercorarium

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997

C:Accession: S12021

R:Jauris, S.; Ruecknagel, K.P.; Schwarz, W.H.; Kratzsch, P.; Bronnenmeier, K.; Staendenbe

Mol. Gen. Genet. 223, 258-267, 1990

A:Title: Sequence analysis of the Clostridium stercorarium celz gene encoding a thermoac

A:Reference number: S12021; MUID:1066838

A:Accession: S12021

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-986 &lt;JAU&gt;

Query Match 14.1% Score 188; DB 2; Length 986;  
Best Local Similarity 27.6%; Pred. No. 1.4e-07;

Matches 47; Conservative 23; Mismatches 58; Indels 36; Gaps 6;

QY 1 MSVEFYNSKSAQNSTIPRIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHAGAL 60

Db 840 IQIQMFNGNTSDKNGIMPRRLNTGTPIRLSDVKIRYYITIDGKDNFMCDSV- 898

QY 61 LGNSYVDNTSKYVIANFVKETASPTSTYDYLDPSSHMRGC--LQSSLLIISVFLVGAQN 118

Db 899 -----GNNITGTFRV-MAEPKEGADYLETGFTDAGYLOPQOSIEVO-----NR 943

QY 119 FSR-----GEVGAATGAVVGAGQLFGKSGRVSMAIGAVLGL 159

Db 944 FSKADMTDIOTNDISFTNTSYG-----SNDRIYVIGVLSGI 984

## RESULT 14

T31085

xylanase - Caldiceilus disruptor sp.

C:Species: Caldiceilus disruptor sp.

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T31085

R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.

submitted to the EMBL data library, December 1997

A:Description: Family 10 and 11 xylanase genes from Caldiceilus disruptor sp. R69B.1.

A:Reference number: Z20972

A:Accession: T31085

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1779 &lt;MOR&gt;

A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1

C:Genetics:

A:Note: xync

Query Match 11.3% Score 151.5; DB 2; Length 1779;  
Best Local Similarity 33.6%; Pred. No. 0.00032;

Matches 51; Conservative 15; Mismatches 35; Indels 51; Gaps 6;

QY 1 MSVEFYNSKSAQNSTIPRIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHA--G 58

Db 1114 LKVLKNNETASGSIKPKRIVGSSVDLSVKIRIYTYVDGDKPQSAVCDMAOIG 1173

QY 59 ALLGNSYVDNTSKYVIANFVKETASPTSTYDYLDPSSHMRGCLQSSLLIISVFLVGAQN 118

Db 1174 A-----SNWTFNFVKLT-SGVSGADYLL----- 1195

QY 119 FSRQEVGAATGAVVGAGQLF-GKSGRVSMA 149

Db 1196 -----EVGFSSG-----AGQLQPGKDTGDIY 1217

## RESULT 15

A43802

cellulase (EC 3.2.1.4) / cellulose 1,4-beta-D-glucosidase (EC 3.2.1.91) - Caldocellu

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Caldocellum saccharolyticum 30-Sep-1993 #text\_change 10-Jul-1998

C:Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993

C:Accession: A43802

R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P

Appl. Environ. Microbiol. 56, 3117-3124, 1990

A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophil

A:Reference number: A43802; MUID:91136262

A:Accession: A43802

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: DNA

A:Residues: 1-915 &lt;SAU&gt;

A:Cross-references: EMBL:X13602

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology &lt;SXY&gt;

Query Match 11.0% Score 147; DB 2; Length 915;  
Best Local Similarity 23.2%; Pred. No. 0.00034;

Matches 60; Conservative 35; Mismatches 64; Indels 100; Gaps 11;

QY 1 MSVEFYNSKSAQNSTIPRIKITNTSDSLNDVKRYRYTSDGTGQTFMCDHA--G 58

Db 351 IKVLYANKETNTSTNTTTPMLKVNSSSIDLSKRTIRYTYVDGERAQSAVSMDAIG 410

QY 59 ALLGNSYVDNTSKYVIANFVKETASPTSTYDYLDPSSHMRGCLQSSLLIISVFLVGAQN 118

Db 411 A-----SNWTFNFVK-LSSVSGADYLL----- 432

QY 119 FSRQEVGAATGAVVGAGQLF-GKSGRVSMAIGAVLGLGKSGISQSDQDOKILN 177

Db 433 -----EIGFKSG-----AGQLQPGKDTGEI-----QIRFN 457

QY 178 OSLEKVKAGQYTRMRNPDTGNSYSVEPYRYQRYNKKOEROOYCFPOKAMINGOKEI 237

Db 458 KS-----DMSNYNQGNDMSV--LQSMSTYGENEKVTATY---DGLVWQGE--- 498

QY 238 YGTACPPQPDGRWQVISTEK 256

Db 499 -----PSTNDDMVLYSGNK 512

Search completed: October 27, 2002, 10:58:43  
Job time : 19.8014 secs

**THIS PAGE BLANK (USPTO)**